THE AVERAGED CONTROL SYSTEM OF FAST OSCILLATING CONTROL SYSTEMS*

ALEX BOMBRUN† AND JEAN-BAPTISTE POMET‡

Abstract. For control systems that either have an explicit periodic dependence on time or have periodic solutions and small controls, we define an *average control system* that takes into account all possible variations of the control, and prove that its solutions approximate all solutions of the oscillating system as oscillations go faster.

The dimension of its velocity set is characterized geometrically. When it is maximum the average system defines a Finsler metric, unfortunately not twice differentiable in general. For minimum time control, this average system allows one to give a rigorous proof that averaging the Hamiltonian given by the maximum principle is a valid approximation.

Key words. Averaging, control systems, small control, optimal control, Finsler geometry.

AMS subject classifications. 34C29, 34H05, 49J15, 93B11, 93C15, 93C70, 53B40

1. Introduction.

We consider either a "fast-oscillating control system" (1):

$$\dot{x} = u_1 X_1(\frac{t}{\varepsilon}, x) + \cdots + u_m X_m(\frac{t}{\varepsilon}, x)\,,\ \ \|u\| \leq 1\,,$$

where all X_i's are 2π-periodic with respect to t/ε, or a "Kepler control system" (47):

$$\dot{\xi} = f_0(\xi) + v_1 f_1(\xi) + \cdots + v_m f_m(\xi)\,,\ \ \|v\| \leq \varepsilon$$

where all solutions of $\dot{\xi} = f_0(\xi)$ are periodic.

Averaging techniques for conservative —periodic or not— ordinary differential equations (ODEs) date back at least to H. Poincaré; see [2, §52] or [23] for recent expositions. Roughly speaking, on a fixed interval, the solutions of $\dot{x} = F(t/\varepsilon, x)$ differ from these of $\dot{x} = \overline{F}(x)$ by a term of order ε, with $\overline{F}$ the average of F with respect to its first argument.

If u or v above is assigned to be a fixed function of state and time (or computed from additional state variables as in $u = \alpha(p, x),\ \dot{p} = g(p, x)$), then these techniques for ODEs can be applied to give an approximation at first order with respect to small ε of the movement of the slow variables. Averaging is usually used in this way in control theory: in vibrational control [19], fast oscillating controls are designed and averaging techniques allows analysis and proof of stability; in the same way, it solves stability and path planning questions in control of mechanical systems, see for instance [8]; in [12, §5], high frequency control is used to approach a non-flat system by a flat one; one may also mention many applications to control [21, 18, 20] of the work [17] that mimics Lie brackets by highly oscillatory controls along the original vector fields. A common feature to these references is that the use of oscillations "creates" new independent controls used for design. The use of averaging in optimal control of oscillating systems [10, 13, 14, 7] is similar in spirit to the above, but closer to the framework of this paper because oscillations are present in the system instead of being introduced by the control. Very interesting results are obtained applying averaging to

* This work was partly supported by *Thales Alenia Space*, in 2004-2007.
Submitted to SIAM J. Control. Optim., December 5, 2011. Revised September 10, 2012.
†Most of this work was done when this author was with INRIA. Alex.Bombrun@gmail.com.
‡INRIA, B.P. 93, 06902 Sophia Antipolis cedex, France. Email: Jean-Baptiste.Pomet@inria.fr.

the Hamiltonian equations arising from Pontryagin Maximum principle. For instance, in [7], the authors have studied in this way the problem of minimal energy transfer between two elliptic orbits; extremals are the same as those giving the geodesics of a Riemannian metric. Again, averaging introduces "new independent controls": Riemannian geodesics are minimizers of a problem where all velocity directions are allowed whereas the velocity set of the original system at each point had positive codimension. The same averaging computation may be applied to the Hamiltonian differential equation obtained for minimum time, but, since this differential equation is discontinuous, there is no theoretical justification for averaging in that case.

Our contribution is to introduce a different way of averaging that takes into account all possible variations of the control —hence the control strategy can be decided *after* performing averaging— and to prove that it has satisfying regularity properties and is a good first order approximation of the above systems as $\varepsilon \to 0$. This gives, as a side result, a justification of the use of averaging for minimum time in [13, 14]. This procedure also "creates new independent control", i.e. increases the dimension of the velocity set, that we characterize in terms of the original vector fields. When this dimension is maximum, the average system defines a Finsler metric [3] on the manifold, whose geodesics are the limits of minimum time trajectories for the original systems as $\varepsilon \to 0$. This Finsler metric is in general not twice differentiable (hence it is not a Finsler metric in the sense of [3], indeed); we however prove that, at least in the less degenerate case, the Hamiltonian system governing extremals, although it is not locally Lipschitz, generates a flow on the cotangent bundle. Low thrust planar orbit transfer belongs to this less degenerate case.

The average control system may be used for other purposes than optimal control, for instance [4] designs a Lyapunov function for feedback control in the average system and uses it for the oscillating systems; indeed the present work was developed out of comparing feedback control based on a priori chosen Lyapunov functions with minimum time control for low thrust orbital transfer.

Preliminary versions of this work can be found in [5, 4].

Organization of the paper. The construction and results are developed for "fast-oscillating control system" in §3 and then transferred in §4 to "Kepler control systems", and applied to minimum time orbit transfer in the planar 2-body problem in §5. One proof is postponed until the appendix, that comes after a short conclusion (§6).

2. Notations and conventions.

2.1. M is a smooth connected manifold of dimension n; its tangent and cotangent bundles are denoted by $\mathrm{T}M$ and T^*M. One may assume for simplicity $M = \mathbb{R}^n$, $\mathrm{T}M = \mathbb{R}^n \times \mathbb{R}^n$, $\mathrm{T}^*M = \mathbb{R}^n \times (\mathbb{R}^n)^*$, and, for $x \in M$, $\mathrm{T}_xM = \mathbb{R}^n$, $\mathrm{T}_x^*M = (\mathbb{R}^n)^*$.

For $v \in \mathrm{T}_xM$, $p \in \mathrm{T}_x^*M$ (or any v, p taken in a vector space and its dual), we denote by $\langle p, v\rangle$ (rather than $p(v)$) their duality product.

2.2. If E is a subset of a vector space V, then $E^\perp$ is its annihilator, the vector subspace of its dual V^* made of all p's such that $\langle p, v\rangle = 0$ for all v in E.

2.3. We assume that M is endowed with an arbitrary Riemannian distance d. If $M = \mathbb{R}^n$, just choose the canonical Euclidean distance.

In local coordinates, $\|.\|$ and $(.\,|.)$ stand for the canonical Euclidean norm and scalar product. On a compact coordinate chart, $k_1\|x - y\| \leq d(x, y) \leq k_2\|x - y\|$ for some positive k_1, k_2 (Lipschitz equivalence). We also denote operator norms by $\|.\|$.

2.4. S^1 is $\mathbb{R}/2\pi\mathbb{Z}$. For θ in S^1 (an angle), we denote by $\mu(\theta)$ the unique real number in $[0, 2\pi)$ such that $\mu(\theta) \equiv \theta \mod 2\pi$. For a real number $s \in \mathbb{R}$, we denote the angle it represents by $s \bmod 2\pi$; it belongs to the quotient S^1.

Maps $S^1 \to E$ (arbitrary set) are identified with 2π-periodic maps $\mathbb{R} \to E$. For instance, if f is such a map $S^1 \to E$ and $\tau \in \mathbb{R}$, we write $f(\tau)$ instead of $f(\tau \bmod 2\pi)$; the average of f is denoted by $\frac{1}{2\pi}\int_0^{2\pi} f(\theta)\mathrm{d}\theta$, or $\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)\mathrm{d}\theta$, or $\frac{1}{2\pi}\int_{\theta\in S^1} f(\theta)\mathrm{d}\theta$; one identifies $L^p(S^1,\mathbb{R}^m)$ with the subset of $L^p(\mathbb{R},\mathbb{R}^m)$ made of 2π-periodic functions.

2.5. The Euclidean norm in $\mathbb{R}^m$ or $(\mathbb{R}^m)^*$ is denoted by $\|.\|$, and the ball of radius one centered at the origin by B^m. We view an element of $\mathbb{R}^m$ as $m \times 1$ matrix (column) of real numbers and an element of $(\mathbb{R}^m)^*$ as a $1 \times m$ matrix (line); transposition, denoted $.^\top$, sends $\mathbb{R}^m$ to $(\mathbb{R}^m)^*$ and vice-versa.

3. Fast oscillating control systems. We call *fast oscillating control system* on M a family of non-autonomous systems, linear in the control $u \in \mathbb{R}^m$:

$$\dot{x} = \mathcal{G}(\frac{t}{\varepsilon}, x)\, u = \sum_{i=1}^{m} \mathcal{G}_i(\frac{t}{\varepsilon}, x)\, u_i\,,\ \|u\| \leq 1 \tag{1}$$

indexed by a positive number ε. Each $\mathcal{G}_i$ is a smooth "periodic time-varying" vector field: $\mathcal{G}_i \in C^\infty(S^1 \times M, \mathrm{T}M)$. An admissible control is a measurable $u(.) : [0,T] \to B^m$ for some $T > 0$. For a given control $u(.)$ and initial condition $x(0)$, there is a unique solution $x(.)$, defined either on $[0,T]$ or only on a maximal interval $[0,T')$, $T' < T$.

Remark 3.1. Apart from being a notation defined by the double equality in (1), $\mathcal{G}(\theta,x)$ defines a linear map $\mathbb{R}^m \to \mathrm{T}_xM$ that sends $(u_1,\ldots,u_m)^\top$ to $\sum_{i=1}^m \mathcal{G}_i(\frac{t}{\varepsilon},x)u_i$.

3.1. Average control system of fast oscillating control systems. Define the map $\overline{\mathcal{G}} : M \times L^\infty([0,2\pi],\mathbb{R}^m) \to \mathrm{T}M$ by

$$\overline{\mathcal{G}}(x,\mathcal{U}) = \frac{1}{2\pi}\int_0^{2\pi} \mathcal{G}(\theta,x)\,\mathcal{U}(\theta)\,\mathrm{d}\theta\ . \tag{2}$$

It allows one to define, for all $x \in M$, the subset $\mathcal{E}(x) \subset \mathrm{T}_xM$ by:

$$\mathcal{E}(x) = \left\{\overline{\mathcal{G}}(x,\mathcal{U}),\ \mathcal{U} \in L^\infty\left([0,2\pi],\mathbb{R}^m\right),\ \|\mathcal{U}\|_\infty \leq 1\right\} \subset T_xM\ , \tag{3}$$

and the *average control system* of (1) as follows; we explain in the following section to what extent it is the limit of (1) as $\varepsilon \to 0$.

Definition 3.2. *The average control system of* (1) *is the differential inclusion*

$$\dot{x} \in \mathcal{E}(x). \tag{4}$$

A solution of (4) is an absolutely continuous $x(.) : [0,T] \to M$ such that $\dot{x}(t) \in \mathcal{E}(x(t))$ for almost all t.

Proposition 3.3. *For all x in M, $\mathcal{E}(x)$ is convex, compact and symmetric with respect to the origin.*

Proof. It is closed, convex and symmetric because it is the image of the unit ball of $L^\infty\left(S^1,\mathbb{R}^m\right)$ by a linear map; it is compact because $\mathcal{G}(x,.)$ is bounded on S^1. ☐

Further characterizations of $\mathcal{E}(x)$ use the map $H : \mathrm{T}^*M \to [0,+\infty)$ defined by

$$H(x,p) = \frac{1}{2\pi}\int_0^{2\pi} \|\langle p, \mathcal{G}(\theta,x)\rangle\|\mathrm{d}\theta \tag{5}$$

where the Euclidean norm is used according to §2.5 and, for each (θ,x),

$$\langle p, \mathcal{G}(\theta,x)\rangle = (\langle p, \mathcal{G}_1(\theta,x)\rangle, \ldots, \langle p, \mathcal{G}_m(\theta,x)\rangle) \in (\mathbb{R}^m)^*\ . \tag{6}$$

PROPOSITION 3.4. *For all $(x,p) \in T^*M$, one has, with H defined in* (5),

$$\mathcal{E}(x) = \Big\{ v \in \mathrm{T}_x M\,,\ \sup_{\substack{p\in \mathrm{T}^*_x M\\ H(x,p)\le 1}} \langle p, v\rangle \ \le\ 1 \Big\}\,, \tag{7}$$

$$H(x,p) \ = \sup_{v\in\mathcal{E}(x)} \langle p,v\rangle \ = \sup_{\mathcal{U}\in L^\infty(S^1,\mathbb{R}^m),\ \|\mathcal{U}\|_\infty\le 1} \langle p, \overline{\mathcal{G}}(x,\mathcal{U})\rangle \ = \ \langle p, \overline{\mathcal{G}}(x,\mathcal{U}^*_{p,x})\rangle\,, \tag{8}$$

with $\mathcal{U}^*_{p,x} \in L^\infty\left(S^1,\mathbb{R}^m\right)$ *defined by:* $$\mathcal{U}^*_{p,x}(\theta) = \begin{cases} 0 & \text{if } \langle p, \mathcal{G}(\theta,x)\rangle = 0\,, \\ \frac{\langle p,\mathcal{G}(\theta,x)\rangle^\top}{\|\langle p,\mathcal{G}(\theta,x)\rangle\|} & \text{if } \langle p, \mathcal{G}(\theta,x)\rangle \neq 0\,. \end{cases} \tag{9}$$

Proof. The last equality in (8) is a straightforward maximization, the second one comes from the definition (3) of $\mathcal{E}(x)$ and a simple computation yields $H(x,p) = \langle p, \overline{\mathcal{G}}(x,\mathcal{U}^*_{p,x})\rangle$; this proves (8). Being closed and convex, $\mathcal{E}(x)$ is the intersection of all its supporting half-spaces [24, Corollary 1.3.5]; according to (8), this yields the following relation, equivalent to (7): $\mathcal{E}(x) = \bigcap_{p\in\mathrm{T}^*_xM} \{v \in \mathrm{T}_xM\,,\ \langle p,v\rangle \le H(x,p)\}$. ☐

A convenient characterization of solutions of (4)*.* According to Definition 3.2, a solution $x(.)$ is such that, for almost all t, there is $\mathcal{U}(t) \in L^\infty([0,2\pi],\mathbb{R}^m)$ such that $\dot x(t) = \overline{\mathcal{G}}(x(t),\mathcal{U}(t))$; the map $(t,\theta) \mapsto \mathcal{U}(t)(\theta)$ is measurable with respect to θ but bears no regularity with respect to t. It turns out that it may always be chosen jointly measurable with respect to (t,θ) according to the following "measurable selection" result:

PROPOSITION 3.5. *A map $x : [0,T] \to \mathbb{R}^n$ is a solution of the differential inclusion* (4) *if and only if there exists $\widehat u \in L^\infty([0,T]\times S^1,\mathbb{R}^m)$, $\|\widehat u\|_\infty \le 1$ such that*

$$\dot x(t) = \frac{1}{2\pi}\int_0^{2\pi} \mathcal{G}(\theta,x(t))\,\widehat u(t,\theta)\,\mathrm{d}\theta \tag{10}$$

for almost all t in $[0,T]$.

Proof. After possibly partitioning $[0,T]$ into intervals where $\dot x(t)$ remains in the same coordinate chart, we work in coordinates and use a Euclidean norm when useful.

Sufficiency is clear: from Fubini theorem, $\theta \mapsto \widehat u(t,\theta)$ is measurable for almost all t, hence $x(.)$ is a solution of (4). Conversely, let $x(.)$ be a solution of (4): $\dot x(.)$ is measurable and, for almost all t, there exists $\tilde u_t \in L^\infty(S^1,\mathbb{R}^m)$, $\|\tilde u_t\|_\infty \le 1$ such that

$$\dot x(t) = \overline{\mathcal{G}}(x(t),\tilde u_t) = \frac{1}{2\pi}\int_0^{2\pi} \mathcal{G}(s_1,x(t))\tilde u_t(s_1)\mathrm{d}s_1\,. \tag{11}$$

Let $\phi : L^\infty\left([0,T]\times S^1,\mathbb{R}^m\right) \to L^2\left([0,T],\mathbb{R}^n\right)$ be the linear map defined by

$$\phi(u)(t) = \overline{\mathcal{G}}(x(t),u(t,.)) = \frac{1}{2\pi}\int_0^{2\pi} \mathcal{G}(s_1,x(t))\,u(t,s_1)\,\mathrm{d}s_1$$

and $\mathcal{I}$ the image by ϕ of the unit ball of $L^\infty([0,T]\times S^1,\mathbb{R}^m)$. Since, by (11), $\dot x(.)$ is essentially bounded, it is in $L^2\left([0,T],\mathbb{R}^n\right)$; since $\mathcal{I}$ is closed and convex in that Hilbert space, the distance from $\dot x$ to $\mathcal{I}$ is reached for a unique element $\bar\xi \in \mathcal{I}$:

$$\bar\xi = \phi(\bar u)\,,\quad \bar u \in L^\infty\left([0,T]\times S^1,\mathbb{R}^m\right)\,,\quad \|\bar u\|_{L^\infty} \le 1\,.$$

Let us prove by contradiction that $\bar\xi = \dot x$, *i.e.* $\dot x(.) \in \mathcal{I}$; this will end the proof.

If $\dot{x} \neq \bar{\xi}$, one has, for all u in the unit ball of $L^\infty\big([0,T]\times S^1,\mathbb{R}^m\big)$,

$$\big(\dot{x}-\bar{\xi}\,\big|\,\phi(u)-\phi(\bar{u})\big)_{L^2} \;\leq\; 0 \tag{12}$$

with equality only if $\phi(u)=\phi(\bar{u})$. Define $\widehat{u}$ by $\widehat{u}(t,s)=\mathcal{U}^*_{\left(\dot{x}(t)-\bar{\xi}(t)\right)^\top,\,x(t)}(s)$ with $\mathcal{U}^*_{p,x}$ defined by (9); clearly, $\widehat{u}$ is in the unit ball of $L^\infty\big([0,T]\times S^1,\mathbb{R}^m\big)$, and, for all $(t,s)\in[0,T]\times S^1$ and all $\mathbf{u}\in\mathbb{R}^m$,

$$\|\mathbf{u}\|\leq 1 \quad\Rightarrow\quad \big(\dot{x}(t)-\bar{\xi}(t)\big)^\top\,\mathcal{G}(s,x(t))\;(\widehat{u}(t,s)-\mathbf{u})\geq 0\,, \tag{13}$$

hence $\big(\dot{x}(t)-\bar{\xi}(t)\big)^\top\mathcal{G}(s_1,x(t))\;(\widehat{u}(t,s_1)-\bar{u}(t,s_1))$ is non-negative for almost all (t,s_1) and, since it is the integrand of the left-hand side of (12), it must be zero; hence $\bar{\xi}=\phi(\bar{u})=\phi(\widehat{u})$ and $\bar{\xi}(t)=\overline{\mathcal{G}}(x(t),\widehat{u}(t,.))$ for almost all t.
In (11), $\tilde{u}_t$ satisfies $\|\tilde{u}_t(s_1)\|\leq 1$ for almost all s_1, hence, according to (13),

$$\big(\dot{x}(t)-\bar{\xi}(t)\big)^\top\,\mathcal{G}(s_1,x(t))\;(\widehat{u}(t,s_1)-\tilde{u}_t(s_1))\geq 0\,.$$

Since $\dot{x}(t)=\overline{\mathcal{G}}(x(t),\tilde{u}_t)$, $\bar{\xi}(t)=\overline{\mathcal{G}}(x(t),\widehat{u}(t,.))$, the integration with respect to the variable s_1 yields $-\|\dot{x}(t)-\bar{\xi}(t)\|^2\geq 0$ for almost all t; this contradicts $\dot{x}\neq\bar{\xi}$. □

Remark 3.6. The differential inclusion (4) is equivalent to the "control system"

$$\dot{x}=\overline{\mathcal{G}}(x,\mathcal{U})\,,\quad \mathcal{U}\in L^\infty(S^1,\mathbb{R}^m)\,,\ \|\mathcal{U}\|_\infty\leq 1$$

where, by Proposition 3.5, admissible controls are maps $t\mapsto\mathcal{U}(t)$ such that $\widehat{u}\colon(t,\theta)\mapsto\mathcal{U}(t)(\theta)$ is measurable with respect to (t,θ). Since this "control" is infinite dimensional, and we could not find a representation of the type $\dot{x}=f(x,v)$, $v\in U\subset\mathbb{R}^r$, r finite, we stay with the differential inclusion (4), with $\mathcal{E}(x)$ described by (8) and (5).

3.2. Convergence theorem. The following result relates solutions of the fast oscillating systems as ε tends to zero to solutions of the average system. To our knowledge, this kind of theorem where the control is not chosen prior to averaging has never been stated in the literature.

Theorem 3.7 (Convergence for fast-oscillating control systems).

1. *Let $x_0(.):[0,T]\to M$ be an arbitrary solution of* (4)*. There exist a family of measurable functions $\overline{u}_\varepsilon(.):[0,T]\to B^m$, indexed by $\varepsilon>0$, and positive constants c,ε_0, such that, calling $x_\varepsilon(.)$ the solution of* (1) *with control $u=\overline{u}_\varepsilon(t)$ and initial condition $x_\varepsilon(0)=x_0(0)$, one has: $x_\varepsilon(.)$ is defined on $[0,T]$ for all ε smaller than ε_0 and converges to $x_0(.)$ as $\varepsilon\to 0$, with an error of uniform order ε:*

$$d(x_\varepsilon(t),x_0(t))<c\,\varepsilon\,,\quad t\in[0,T]\,,\ 0<\varepsilon<\varepsilon_0\,. \tag{14}$$

2. *Let $\mathbb{K}$ be a compact subset of M, $(\varepsilon_n)_{n\in\mathbb{N}}$ a decreasing sequence of positive real numbers converging to zero, and, for each n, $x_n(.)\colon[0,T]\to\mathbb{K}$ a solution of (1) with $\varepsilon=\varepsilon_n$ and for some control $u=u_n(t)$, $u_n(.)\in L^\infty([0,T],\mathbb{R}^m)$, $\|u_n(.)\|_\infty\leq 1$. Then the sequence $\big(x_n(.)\big)_{n\in\mathbb{N}}$ is compact for the topology of uniform convergence on $[0,T]$ and any accumulation point is a solution of the average system* (4)*.*

The statement is more complex than the one for ODEs, *e.g.* [2, §52.C], due to underdetermination (choice of control in (1), multi-valued right-hand side in (4)).

Informally, "1" states that any solution of the average system is the limit of solutions of fast oscillating systems with well chosen controls and "2" states that, conversely, any limit of solutions of oscillating systems, with arbitrary controls, is a solution of the average control system. There is an estimate on the error in "1" but not in "2" because some sequences may converge slower than others.

Remark 3.8. We could have considered systems that are affine instead of linear in the control by adding a drift vector field $\mathcal{G}_0(t/\varepsilon, x)$ to (1). One then obtains an average control system where $\mathcal{E}(x)$ is replaced by $\overline{\mathcal{G}}_0(x)+\mathcal{E}(x)$, with $\overline{\mathcal{G}}_0(x) = \frac{1}{2\pi}\int_0^{2\pi}\mathcal{G}_0(\theta,x)\mathrm{d}\theta$. Convergence does hold. We restrain to systems (1) because they naturally occur when dealing with a time-invariant system whose drift has periodic solutions, see §4 below.

In the proof of Theorem 3.7, the following technical lemma is needed.

LEMMA 3.9. *Let $\varepsilon > 0$ and $a < b$ be real numbers and $\widehat{u} : [a-2\pi\varepsilon, b]\times S^1 \to \mathbb{R}^m$ be measurable, one has the following identity (see §2.4 for the notation $\mu(.)$) :*

$$\iint_{\substack{\theta\in S^1\\ a\le s\le b}} \mathcal{G}(\theta, x(s))\ \widehat{u}(s,\theta)\ \mathrm{d}\theta\,\mathrm{d}s = \iint_{\substack{\theta\in S^1\\ a\le s\le b}} \mathcal{G}(\frac{s}{\varepsilon}, x(s))\ \widehat{u}(s+\varepsilon\mu(\theta), \frac{s}{\varepsilon})\ \mathrm{d}\theta\,\mathrm{d}s + \Delta_\varepsilon \tag{15}$$

with
$$\begin{aligned}\Delta_\varepsilon = &\iint_{T^a_\varepsilon} \mathcal{G}(\frac{s}{\varepsilon}, x(s+\varepsilon\mu(\theta)))\,\widehat{u}(s+\varepsilon\mu(\theta), \frac{s}{\varepsilon})\,\mathrm{d}\theta\,\mathrm{d}s\\ &-\iint_{T^b_\varepsilon} \mathcal{G}(\frac{s}{\varepsilon}, x(s+\varepsilon\mu(\theta)))\,\widehat{u}(s+\varepsilon\mu(\theta), \frac{s}{\varepsilon})\,\mathrm{d}\theta\,\mathrm{d}s\\ &+\iint_{\substack{\theta\in S^1\\ a\le s\le b}} \Big[\mathcal{G}(\frac{s}{\varepsilon}, x(s+\varepsilon\mu(\theta))) - \mathcal{G}(\frac{s}{\varepsilon}, x(s))\Big]\,\widehat{u}(s+\varepsilon\mu(\theta), \frac{s}{\varepsilon})\,\mathrm{d}\theta\,\mathrm{d}s\end{aligned} \tag{16}$$

and the set T^a_ε defined by $T^a_\varepsilon = \{(s,\theta),\ \theta\in S^1,\ a-\varepsilon\,\mu(\theta)\le s\le a\}$ and T^b_ε accordingly.

Proof. Thanks to the change of variables $\theta = \tau/\varepsilon \bmod 2\pi$, $s = \tau + \varepsilon\,\mu(\phi)$, with $\mu(\theta)$ as defined in §2.4, the left-hand side of (15) is equal to

$$\iint_{\substack{\phi\in S^1\\ a-\varepsilon\phi\le\tau\le b-\varepsilon\mu(\phi)}} \mathcal{G}(\frac{\tau}{\varepsilon}, x(\tau+\varepsilon\mu(\phi)))\,\widehat{u}(\tau+\varepsilon\mu(\phi), \frac{\tau}{\varepsilon})\,\mathrm{d}\tau\,\mathrm{d}\phi\ .$$

Keeping the names (s,θ) instead of (τ,ϕ), one gets (15), the correcting term Δ_ε coming from the modified domain of integration and argument of x. ◻

Proof of Theorem 3.7, point 1. Consider a solution $x_0 : [0,T] \to M^n$ of (4). By Proposition 3.5 there exists $\widehat{u}_0 \in L^\infty([0,T]\times S^1, \mathbb{R}^m)$, $\|\widehat{u}_0\|_\infty \le 1$ satisfying (10). For $\varepsilon > 0$, define $\overline{u}_\varepsilon(.) \in L^\infty([0,T],\mathbb{R}^m)$ by (see §2.4 for notations):

$$\overline{u}_\varepsilon(t) = \frac{1}{2\pi}\int_0^{2\pi} \widehat{u}_0(t+\varepsilon\,\mu(\theta), \frac{t}{\varepsilon})\,\mathrm{d}\theta\ , \tag{17}$$

where $\widehat{u}_0$ is prolonged by zero outside $[0,T]$: $\widehat{u}_0(t+\varepsilon\,\mu(\theta), \frac{t}{\varepsilon}) = 0$ if $t+\varepsilon\,\mu(\theta) > T$. Let us prove that this construction of $\overline{u}_\varepsilon$ satisfies the two announced properties.

Step 1. Let us first assume that M is an open subset of $\mathbb{R}^n$ and $\mathcal{G}$ is zero outside a compact subset of M. Then $\mathcal{G}(\theta,x)$ is a $n\times m$ matrix for all (θ,x) and, denoting by $\|.\|$ the Euclidean norm for vectors and the operator norm for matrices, there are global constants $\operatorname{Lip}\mathcal{G}$ and $\sup\mathcal{G}$ such that, for all x, x', θ in $M\times M\times S^1$,

$$\|\mathcal{G}(\theta,x)-\mathcal{G}(\theta,x')\| \le \big(\operatorname{Lip}\mathcal{G}\big)\|x-x'\|\ ,\quad \|\mathcal{G}(\theta,x)\| \le \big(\sup\mathcal{G}\big)\ . \tag{18}$$

Let b be a non-negative constant and consider, for each $\varepsilon > 0$, a solution $x_\varepsilon(.)$ of (1) with control $u = \overline{u}_\varepsilon(t)$ and initial condition $x_\varepsilon(0)$ such that

$$\|x_\varepsilon(0) - x_0(0)\| \le b\,\varepsilon\ . \tag{19}$$

In fact $b = 0$ in the theorem but we need a nonzero b in step 2. By definition, expanding $\overline{u}_\varepsilon(s)$ as in (17) and using Lemma 3.9, one has

$$\begin{aligned} x_\varepsilon(t) &= x_\varepsilon(0) + \frac{1}{2\pi}\int_0^t\int_0^{2\pi} \mathcal{G}(\frac{s}{\varepsilon}, x_\varepsilon(s))\, \widehat{u}_0(s-\varepsilon\,\mu(\theta), \frac{s}{\varepsilon})\,\mathrm{d}\theta\,\mathrm{d}s\,, \\ &= x_\varepsilon(0) + \frac{1}{2\pi}\left(\int_0^t\int_0^{2\pi} \mathcal{G}(\theta, x_\varepsilon(s))\,\widehat{u}_0(s,\theta)\,\mathrm{d}\theta\,\mathrm{d}s \;-\; \Delta_\varepsilon\right) \end{aligned} \tag{20}$$

with Δ_ε given by (16), that satisfies $\|\Delta_\varepsilon\| \le 4\pi^2\big(\mathrm{Lip}\,\mathcal{G}\big)\,(1+T\sup\mathcal{G})\,\varepsilon$ because, in particular, $\|\widehat{u}_0\| \le 1$, $|\varepsilon\mu(\theta)| < 2\pi\varepsilon$ and

$$\|\Big(\mathcal{G}(\frac{s}{\varepsilon}, x_\varepsilon(s)) - \mathcal{G}(\frac{s}{\varepsilon}, x_\varepsilon(s+\varepsilon\,\mu(\theta)))\Big)\,\widehat{u}_0(s+\varepsilon\,\mu(\theta), \frac{s}{\varepsilon})\| \le 2\pi\,\big(\mathrm{Lip}\,\mathcal{G}\big)\,\big(\sup\mathcal{G}\big)\,\varepsilon\,.$$

Using (19), (20) the above bound on $\|\Delta_\varepsilon\|$ and the relation

$$x_0(t) = x_0(0) + \frac{1}{2\pi}\int_0^t\int_0^{2\pi} \mathcal{G}(\theta, x_0(s))\,\widehat{u}_0(s,\theta)\,\mathrm{d}\theta\,\mathrm{d}s\,,$$

yields

$$\|x_\varepsilon(t) - x_0(t)\| \le \Big(b + 2\pi\,\big(\mathrm{Lip}\,\mathcal{G}\big)\,\big(1+T\sup\mathcal{G}\big)\Big)\,\varepsilon + \big(\mathrm{Lip}\,\mathcal{G}\big)\int_0^t \|x_\varepsilon(s) - x_0(s)\|\,\mathrm{d}s$$

for all t in $[0,T]$, and finally, using the classical Gronwall lemma,

$$\|x_\varepsilon(t) - x_0(t)\| \le \Big[b + 2\pi\big(\mathrm{Lip}\,\mathcal{G}\big)\,(1+T\sup\mathcal{G})\Big]\; e^{T\,\mathrm{Lip}\,\mathcal{G}}\,\varepsilon \tag{21}$$

for all t in $[0,T]$ and ε in $[0,\varepsilon_0]$. This proves the theorem if M is an open subset of $\mathbb{R}^n$ and $\mathcal{G}$ is zero outside a compact subset, with an explicit constant c corresponding to the distance d defined from the Euclidean norm and with $\varepsilon_0 = +\infty$.

Step 2. General case. Let $x_\varepsilon(.)$ be the solution of (1) with control $u = \overline{u}_\varepsilon(t)$ defined in (17) from $\widehat{u}_0$ and with initial condition $x_\varepsilon(0) = x_0(0)$; it is not necessarily defined on $[0,T]$ but may have a maximum interval of definition $[0,T_\varepsilon)$ with $T_\varepsilon < T$. Let $\widetilde{T} \in [0,T]$ be the supremum of the set of numbers $\tau \in [0,T]$ such that, for some ε_0 and some c, that may depend on τ, the solution $x_\varepsilon(.)$ is defined on $[0,\tau]$ and satisfies $d(x_\varepsilon(t), x_0(0)) < c\,\varepsilon$ for all $t \in [0,\tau]$ and $\varepsilon \in [0,\varepsilon_0]$. Let us prove by contradiction that $\widetilde{T} = T$. This will end the proof of Theorem 3.7, point 1.

Assume $\widetilde{T} < T$, and let

- $\mathcal{O}$ be a coordinate neighborhood of $x_0(\widetilde{T})$,
- $\alpha > 0$ be such that $0 < \widetilde{T} - \alpha < \widetilde{T} + \alpha \le T$ and $x_0([\widetilde{T}-\alpha, \widetilde{T}+\alpha]) \subset \mathcal{O}$,
- $c > 0$, $\varepsilon_0 > 0$ be such that $d(x_\varepsilon(t), x_0(t)) < c\,\varepsilon$ for all $t \in [0, \widetilde{T}-\alpha]$ and $\varepsilon \in [0,\varepsilon_0]$.

Taking ε_0 possibly smaller, one also has $x_\varepsilon(\widetilde{T}-\alpha) \in \mathcal{O}$ for $\varepsilon < \varepsilon_0$. Let $\mathbb{K}$ be a compact neighborhood of $x_0([\widetilde{T}-\alpha, \widetilde{T}+\alpha])$ contained in $\mathcal{O}$, $\mathbb{K}'$ a compact neighborhood of $\mathbb{K}$ contained in $\mathcal{O}$, and $\rho : M \to [0,1]$ a smooth map, zero outside $\mathbb{K}'$ and constant equal to 1 in $\mathbb{K}$. Defining $\mathcal{G}_\rho$ by $\mathcal{G}_\rho(\theta,x) = \rho(x)\mathcal{G}(\theta,x)$, let us apply Step 1 in coordinates in $\mathcal{O}$, with $\mathcal{G}_\rho$ instead of $\mathcal{G}$ and $[\widetilde{T}-\alpha, \widetilde{T}+\alpha]$ instead of $[0,T]$. Call x_0^ρ (resp. x_ε^ρ, $\varepsilon > 0$) the solution of (10) (resp. of (1) with control $u = \overline{u}_\varepsilon(t)$), replacing $\mathcal{G}$ by $\mathcal{G}_\rho$, with initial condition $x_\varepsilon^\rho(\widetilde{T}-\alpha) = x_\varepsilon(\widetilde{T}-\alpha)$, $\varepsilon \ge 0$. One clearly has, as in (19), $\|x_\varepsilon^\rho(\widetilde{T}-\alpha) - x_0^\rho(\widetilde{T}-\alpha)\| < b\,\varepsilon$ with b deduced from c via Lipschitz equivalence of the

distance d and the one induced by the norm in coordinates (see §2.3); then Step 1 provides $\varepsilon_0' > 0$ and, by (21), the inequality

$$\|x_\varepsilon^\rho(t) - x_0^\rho(t)\| \le \left[b + 2\pi\big(\mathrm{Lip}\,\mathcal{G}_\rho\big)\,(1 + 2\alpha \sup \mathcal{G}_\rho)\right]\, e^{2\alpha\,\mathrm{Lip}\,\mathcal{G}_\rho}\,\varepsilon\;, \tag{22}$$

valid for $t \in [\widetilde{T} - \alpha, \widetilde{T} + \alpha]$ and $\varepsilon \in [0, \varepsilon_0']$. Possibly choosing a smaller ε_0', this implies that $x_\varepsilon([\widetilde{T} - \alpha, \widetilde{T} + \alpha]) \subset \mathbb{K}$ for $\varepsilon < \varepsilon_0'$; since $\mathcal{G}$ that coincides with $\mathcal{G}_\rho$ in $\mathbb{K}$, the conclusions holds for x_ε and $\mathcal{G}$ as well as x_ε^ρ and $\mathcal{G}_\rho$ if ε is no larger than ε_0'. We have shown that, for all $\varepsilon < \varepsilon_0'$, the solution x_ε is defined on $[0, \widetilde{T} + \alpha]$ and, with c' deduced from c on time time-interval $[0, \widetilde{T} - \alpha]$ and from the explicit constant in (22) and Lipschitz equivalence of d and the Euclidean distance on time time-interval $[\widetilde{T} - \alpha, \widetilde{T} + \alpha]$, it satisfies $d(x_\varepsilon(t) - x_0(t)) \le c'\varepsilon$ for t in $[0, \widetilde{T} + \alpha]$. This contradicts the definition of $\widetilde{T}$. ◻

Proof of Theorem 3.7, point 2. Since $\mathcal{G}$ is bounded on $S^1 \times \mathbb{K}$ (one may cover $\mathbb{K}$ with a finite number of coordinate charts and define this bound in coordinates), the maps $x_n(.)$ have a common Lipschitz constant and the sequence $(x_n(.))$ is equi-continuous, hence compact by Ascoli-Arzela Theorem: one may extract a uniformly convergent sub-sequence. Still denoting by $(x_n(.))_{n\in\mathbb{N}}$ such a converging sub-sequence and by $x^*(.)$ its (uniform) limit, we need to prove that this limit is a solution of (4).

Define, for each n, $\widehat{u}_n : [0, T] \times S^1 \to \mathbb{R}^m$ by

$$\widehat{u}_n(t, \theta) = u_n(\,\beta_n(t, \theta)\,)\,, \tag{23}$$

where $u_n(.) \in L^\infty([0, T], \mathbb{R}^m)$ is associated to $x_n(.)$ according to the assumption of the theorem and where the map $\beta_n : [0, T] \times S^1 \to \mathbb{R}$ is defined by

$$t - 2\pi\varepsilon_n < \beta_n(t, \theta) \le t\;, \quad \frac{\beta_n(t, \theta)}{\varepsilon_n} \equiv \theta \text{ modulo } 2\pi\;. \tag{24}$$

Clearly $\widehat{u}_n$ is in $L^\infty([0, T] \times S^1, \mathbb{R}^m)$ and $\|\widehat{u}_n\|_\infty \le 1$. Hence, after possibly extracting a sub-sequence, $(\widehat{u}_n)$ converges in the weak-$*$ topology to some $\widehat{u}^*$. Let us prove that, for almost all $t \in [0, T]$,

$$\dot{x}^*(t) = \frac{1}{2\pi}\int_0^{2\pi} \mathcal{G}(\theta, x^*(t))\, \widehat{u}^*(t, \theta)\mathrm{d}\theta\;. \tag{25}$$

Let $\widetilde{T} \in [0, T]$ be the supremum of the set of numbers $\tau \in [0, T]$ such that this is true for almost all t in $[0, \tau]$, and let us prove by contradiction that $\widetilde{T} = T$.

Assume $\widetilde{T} < T$, and let $\mathcal{O}$ be a coordinate neighborhood of $x_0(\widetilde{T})$ and α be such that $0 < \widetilde{T} - \alpha < \widetilde{T} + \alpha \le T$ and $x_0([\widetilde{T} - \alpha, \widetilde{T} + \alpha]) \subset \mathcal{O}$. Uniform convergence implies $x_n([\widetilde{T} - \alpha, \widetilde{T} + \alpha]) \subset \mathcal{O}$ for n large enough and then, in coordinates, for $t \in [\widetilde{T} - \alpha, \widetilde{T} + \alpha]$,

$$x_n(t) - x_n(\widetilde{T} - \alpha) = \int_{\widetilde{T}-\alpha}^{t} \mathcal{G}(\frac{s}{\varepsilon_n}, x_n(s))\, u_n(s)\,\mathrm{d}s\,. \tag{26}$$

From (24), one has $\beta_n(s + \varepsilon_n\theta, \frac{s}{\varepsilon_n}) = s$, hence, from (23), $\widehat{u}_n(s + \varepsilon_n\theta, \frac{s}{\varepsilon_n}) = u_n(s)$ for all $\theta \in S^1$, $s \in \mathbb{R}$; using this in Lemma 3.9, one has

$$\tfrac{1}{2\pi}\iint_{\substack{\widetilde{T}-\alpha \le s \le t\\ \theta \in S^1}} \mathcal{G}(\theta, x_n(s))\widehat{u}_n(s, \theta)\,\mathrm{d}\theta\,\mathrm{d}s = \tfrac{1}{2\pi}\iint_{\substack{\widetilde{T}-\alpha \le s \le t\\ \theta \in S^1}} \mathcal{G}(\frac{s}{\varepsilon_n}, x_n(s))\, u_n(s)\,\mathrm{d}\theta\,\mathrm{d}s\; + \Delta_{\varepsilon_n}\;.$$

Since the integral in the right-hand side —whose integrand does not depend on θ— is equal to the right-hand side of (26), one gets, using uniform convergence of x_n to x^*, weak convergence of $\widehat{u}_n$ to $\widehat{u}^*$ and convergence of Δ_{ε_n} to zero,

$$x^*(t) - x^*(\widetilde{T}-\alpha) = \frac{1}{2\pi}\int_{\widetilde{T}-\alpha}^{t}\int_0^{2\pi} \mathcal{G}(\theta, x^*(s))\widehat{u}^*(s,\theta)\,\mathrm{d}\theta\,\mathrm{d}s\ ,$$

for t in $[\widetilde{T}-\alpha, \widetilde{T}+\alpha]$, and finally that (25) hold for almost all t in $[0, \widetilde{T}+\alpha]$, thus contradicting the definition of $\widetilde{T}$. ◻

3.3. Dimension of the velocity set $\mathcal{E}(x)$. Recall that, for a convex subset C of a linear space, containing the origin, its linear hull is the smallest linear subspace that contains C, the interior of C *in its linear hull* is always nonempty, $\dim C$ is the dimension of this linear hull.

Viewing $\frac{\partial^j \mathcal{G}}{\partial\theta^j}(\theta,x)$ as a linear map $\mathbb{R}^m \to \mathrm{T}_x M$ (see Remark 3.1), and Σ denoting a sum of linear subspaces of T_xM, define the integer $r(\theta,x)$ by:

$$r(\theta,x) = \dim\Big(\sum_{j\in\mathbb{N}} \mathrm{Range}\,\frac{\partial^j \mathcal{G}}{\partial\theta^j}(\theta,x)\Big)\ . \tag{27}$$

It is also the rank of the collection of vectors $\frac{\partial^j \mathcal{G}_i}{\partial\theta^j}(\theta,x) \in \mathrm{T}_xM$, $1\le i\le m$, $j\ge 0$.

In the following proposition, and it is the sole place where this property is used, "system (1) is real analytic with respect to θ" means that the vector fields $\mathcal{G}_i$ are real analytic with respect to θ for fixed x (while being smooth with respect to (θ,x)).

Proposition 3.10.

1. *The linear hull of $\mathcal{E}(x)$ satisfies the following two properties for all x in M, where the inclusion (29) is an equality if (1) is real analytic with respect to θ:*

$$\text{Linear hull}\,\mathcal{E}(x) \ = \ \sum_{\theta\in S^1} \mathrm{Range}\,\mathcal{G}(\theta,x)\ , \tag{28}$$

$$\text{Linear hull}\,\mathcal{E}(x) \ \supset \ \sum_{j\in\mathbb{N}} \mathrm{Range}\,\frac{\partial^j \mathcal{G}}{\partial\theta^j}(\theta,x) \qquad \textit{for all } \theta\in S^1\ . \tag{29}$$

2. *If $r(\theta,x)=n$ for at least one θ in S^1, then $\mathcal{E}(x)$ has a nonempty interior in T_xM, i.e.* $\dim\mathcal{E}(x)=n$.

3. *If the system (1) is real analytic with respect to θ, then $r(\theta,x)$ does not depend on θ and $r(\theta,x)=\dim\mathcal{E}(x)$.*

Proof. If p is in $\mathrm{Range}\,\mathcal{G}(\theta,x)^{\perp}$ for all θ, then any $v=\overline{\mathcal{G}}(x,\mathcal{U})$ in $\mathcal{E}(x)$ satisfies $\langle p,v\rangle = 0$ because $\langle p, \mathcal{G}(\theta,x)\mathcal{U}(\theta)\rangle$ is identically zero on $[0,2\pi]$. Conversely, let p be in $\mathcal{E}(x)^{\perp}$, and consider $v=\overline{\mathcal{G}}(x,\mathcal{U}^*_{p,x})\in\mathcal{E}(x)$; then $\langle p,v\rangle=0$ implies $\langle p,\mathcal{G}(\theta,x)\rangle = 0$, i.e. $p\in\mathrm{Range}\,\mathcal{G}(\theta,x)^{\perp}$ for all θ. This proves the identity (28) between subspaces of T_xM through the following one on their annihilators in T_x^*M: $\mathcal{E}(x)^{\perp} = \bigcap_{\theta\in S^1}(\mathrm{Range}\,\mathcal{G}(\theta,x))^{\perp}$. Similarly, and using (28), the inclusion implies (29): $\bigcap_{\phi\in S^1}(\mathrm{Range}\,\mathcal{G}(\phi,x))^{\perp}\subset\bigcap_{j\in\mathbb{N}}\big(\mathrm{Range}\,\frac{\partial^j\mathcal{G}}{\partial\theta^j}(\theta,x)\big)^{\perp}$. If p is in $\mathrm{Range}\,\mathcal{G}(\phi,x)^{\perp}$ for all ϕ, differentiating $\langle p\,,\mathcal{G}(\phi,x)\rangle=0$ with respect to ϕ an arbitrary number of times yields $\big\langle p\,,\partial^j\mathcal{G}/\partial\phi^j(x,\phi)\big\rangle=0$, $j\in\mathbb{N}$; this proves, with $\phi=\theta$, the above inclusion, hence (29). To prove the converse in the real analytic case, fix $\theta\in S^1$ and $p\in\bigcap_{j\in\mathbb{N}}\frac{\partial^j\mathcal{G}}{\partial\theta^j}(\theta,x)^{\perp}$, and consider the real analytic mapping $S^1\to(\mathbb{R}^m)^*$, $\phi\mapsto\langle p\,,\mathcal{G}(\phi,x)\rangle$; the assumption on p implies that this map vanishes for $\phi=\theta$, as well as its derivatives at all

orders, hence is identically zero: $p \in \bigcap_{\phi\in S^1} \left(\mathrm{Range}\,\mathcal{G}(\phi,x)\right)^{\perp}$. This ends the proof of Point 1. Point 2 is an easy consequence and Point 3 is classical. ☐

3.4. Further properties in the full rank case. We now assume that the mapping $\mathcal{G}$ in (1) is such that the rank $r(\theta,x)$ defined by (27) is maximal:

$$r(\theta,x)=n \;\text{ for all } x \text{ in } M \text{ and } \theta \text{ in } S^1. \tag{30}$$

3.4.1. Controllability. This condition is strongly related to controllability of the linear approximation of (1) around equilibria, i.e. around the solutions where x constant and u identically zero. Indeed, the linear approximation of the time-varying nonlinear system (take $\varepsilon=1$ in (1)):

$$\dot{x}=\mathcal{G}(t,x)u \tag{31}$$

around the equilibrium $x=x_1$ is the time-varying linear system $\dot{\xi}=\mathcal{G}(t,x_1)u$; according to [16, p.614], it is "controllable with impulsive controls at any time" if and only if $r(t,x_1)=n$ for all t. If this is true at all points x_1 then all end-point mappings are submersions around zero controls; we shall need the following more precise result:

Proposition 3.11. *Assume that (30) holds.*

1. *For all $x_1\in M$ and $T>0$, there exist a coordinate neighborhood $\mathcal{W}$ of x_1 (the ball $\mathcal{B}$ below refers to the Euclidean norm in these coordinates), positive constants α_0, c_3, and, for all $y\in\mathcal{W}$, a smooth map $\chi_y:\mathcal{B}(y,\alpha_0)\to L^\infty([0,T],\mathbb{R}^m)$ with Lipschitz constant c_3, which is a right inverse of the end-point mapping of (31) on $[0,T]$ starting from y, i.e. for all $y_f\in\mathcal{B}(y,\alpha_0)$, the control $\chi_y(y_f):[0,T]\to\mathbb{R}^m$ is such that the solution of $\dot{x}=\mathcal{G}(t,x)\chi_y(y_f)(t)$, $x(0)=y$ satisfies $x(T)=y_f$.*
2. *For all $\varepsilon>0$, the system (1) is fully controllable, i.e. there exists, for any $\varepsilon>0$ and any two point x_0,x_1 in M, a time T and a measurable control $u:[0,T]\to B^m$ such that the solution of (1) with $x(0)=x_0$ satisfies $x(T)=x_1$.*

Proof. Let $E_y:L^\infty([0,T],\mathbb{R}^m)\to M$ be the end-point mapping with starting point y. Condition (30) implies that the derivative of E_{x_1} at the zero control has rank n; hence there exists an n-dimensional subspace V of $L^\infty([0,T],\mathbb{R}^m)$ such that the restriction of E_{x_1}, and hence of E_y for y close enough, to V is a local diffeomorphism at zero; the χ_y's are the local inverses of these local diffeomorphisms; they depend smoothly on y, hence the common α_0 and c_3 in Point 1.

This implies that the reachable set from any point at any positive or negative time contains a neighborhood of this point; a classical argument then tells us that the reachable set from a point x_0 is M, assumed to be connected, for it is both open (obvious) and closed (if $\bar{x}$ is in the closure of the reachable set, some points in the reachable set can be reached in negative time, hence $\bar{x}$ can be reached from x_0). ☐

Let us now turn to the average system (4). From $H:\mathrm{T}^*M\to[0,+\infty)$ defined by (5), we define $N:\mathrm{T}M\to[0,+\infty]$ by

$$N(x,v)=\max_{p\in\mathrm{T}_x^*M,\;H(x,p)\le 1}\langle p,v\rangle\ . \tag{32}$$

Proposition 3.12. *Assume the rank condition (30).*

1. *For all $x\in M$, $H(x,.)$ defines a norm on the cotangent space T_x^*M, its dual norm on the tangent space T_xM is $N(x,.)$, and $\mathcal{E}(x)$ is the unit ball for $N(x,.)$, i.e. $\mathcal{E}(x)=\{v\in\mathrm{T}_xM,\ N(x,v)\le 1\}$.*
2. *System (4) is fully controllable, i.e. there exists, for any points x_0,x_1 in M, a time T and a solution $x(.):[0,T]\to M$ of (4) such that $x(0)=x_0$, $x(T)=x_1$.*

Proof. From (5), $H(x,p)=0$ implies $\langle p, \mathcal{G}(\theta,x)\rangle = 0$ for all θ and, differentiating with respect to θ and using (30), this implies $p=0$; this makes $p \mapsto H(x,p)$ a norm for other properties are straightforward. Hence N given by (32) is finite for any (x,v) and it is, by definition, the dual norm of $H(x,.)$; $\mathcal{E}(x)$ is its unit ball by (7) in Proposition 3.4. To prove Point 2, take a continuously differentiable curve $\gamma : [0,1] \to M$ such that $\gamma(0)=x_0$ and $\gamma(1)=x_1$ and $\sigma : [0,T] \to [0,1]$ for some $T>0$, differentiable, such that

$$t \geq \int_0^{\sigma(t)} N(\gamma(s), \frac{\mathrm{d}\gamma}{\mathrm{d}s}(s))\mathrm{d}s$$

(N and H are obviously continuous), then $t \mapsto x(t) = \gamma(\sigma(t))$ is a solution of (4) such that $x(0)=x_0$ and $x(T)=x_1$. ◻

3.4.2. On the differentiability of H. It is clear that H, given by (5), is as smooth as $\mathcal{G}$ on $\mathrm{T}^*M \setminus \widetilde{\mathcal{Z}}$ (where the square root does not vanish) with

$$\widetilde{\mathcal{Z}} = \{(x,p) \in \mathrm{T}^*M,\, \exists \theta \in S^1,\, \langle p, \mathcal{G}(\theta,x)\rangle = 0\}\ . \tag{33}$$

Unfortunately $\widetilde{\mathcal{Z}}$ is not empty: it is, generically (away from $\{p=0\}$) a $2n-m+1$ dimensional submanifold of T^*M. Let us give a first statement valid even on $\widetilde{\mathcal{Z}}$. We state it for $H^2 : (x,p) \mapsto H(x,p)^2$, because H itself, homogeneous of degree 1 with respect to p, cannot be differentiable on $\{p=0\}$, that coincides with $\{H(x,p)=0\}$ by Proposition 3.12 item 1.

THEOREM 3.13. *If condition (30) holds, H^2 is continuously differentiable.*

We first prove a more generic property. In the following proposition, the notations are independent from the rest of the paper:

PROPOSITION 3.14. *Let d be a positive integer, O^{d} an open subset of $\mathbb{R}^{\mathsf{d}}$, V: $S^1 \times O^{\mathsf{d}} \to \mathbb{R}^m$ a smooth map (C^∞), $\widetilde{\mathcal{Z}}$ the subset of O^{d} where V vanishes for some θ and H: $O^{\mathsf{d}} \to [0,+\infty)$ the average of the norm of V:*

$$\mathsf{H}(X) = \frac{1}{2\pi}\int_0^{2\pi} \|\mathsf{V}(\theta,X)\|\,\mathrm{d}\theta\ , \quad \widetilde{\mathcal{Z}} = \{X \in O^{\mathsf{d}},\, \exists\theta \in S^1,\, \mathsf{V}(\theta,X)=0\}\ . \tag{34}$$

Assume:

for all X in O^{d}, the set $\{\theta \in S^1\ ,\ \mathsf{V}(\theta,X)=0\}$ has measure zero in S^1. (35)

Then H is continuously differentiable and, for all X,

$$\mathrm{d}\mathsf{H}(X).h = \frac{1}{2\pi}\int_0^{2\pi} \left(\frac{\partial \mathsf{V}}{\partial X}(\theta,X).h \,\middle|\, \frac{\mathsf{V}(\theta,X)}{\|\mathsf{V}(\theta,X)\|}\right) \mathrm{d}\theta\ . \tag{36}$$

Proof. The integral in (36) is well defined (its integrand is bounded) and, by (35) and Lebesgue convergence theorem, it is continuous with respect to X and h. Let us prove that this $\mathrm{d}\mathsf{H}$ *is* the derivative of H. Since V is smooth, one has

$$\|\mathsf{V}(\theta,X+h) - \mathsf{V}(\theta,X) - \frac{\partial\mathsf{V}}{\partial X}(\theta,X).h\| \leq k\,\|h\|^2\,, \tag{37}$$

where $\frac{\partial\mathsf{V}}{\partial X}(\theta,X)$ is smooth with respect to (θ,X) and k is some local constant. Now, assuming $\mathsf{V}(\theta,X)\neq 0$, one has

$$\|\mathsf{V}(\theta,X+h)\| - \|\mathsf{V}(\theta,X)\| = \left(\mathsf{V}(\theta,X+h) - \mathsf{V}(\theta,X) \,\middle|\, \frac{\mathsf{V}(\theta,X)}{\|\mathsf{V}(\theta,X)\|}\right) \\ + a(\theta,X,h)\,\frac{\|\mathsf{V}(\theta,X+h)-\mathsf{V}(\theta,X)\|^2}{\|\mathsf{V}(\theta,X)\| + \|\mathsf{V}(\theta,X+h)\|}$$

with $|a(\theta,X,h)|\leq 2$. Hence, from (37) and (36), one has, for some local constant k',

$$\frac{\|\mathsf{H}(X+h)-\mathsf{H}(X)-\mathrm{d}\mathsf{H}(X).h\|}{\|h\|}\leq\frac{k'}{2\pi}\int_0^{2\pi}\left(\|h\|+\frac{\|\mathsf{V}(\theta,X+h)-\mathsf{V}(\theta,X)\|}{\|\mathsf{V}(\theta,X)\|+\|\mathsf{V}(\theta,X+h)\|}\right)\mathrm{d}\theta$$

for $\|h\|$ small enough. For fixed X and $h\to 0$, the integrand in the right-hand side is bounded by $1+\|h\|$ and converges to zero for θ outside the set $\{\theta\in S^1,\ \mathsf{V}(\theta,X)=0\}$: by (35) and Lebesgue convergence theorem, the right-hand side tends to zero. ◻

Proof of Theorem 3.13. We operate in coordinates for this is anyway a local property. Apply Proposition 3.14 with $\mathsf{d}=2n$, O^{d} a neighborhood of a point where $p\neq 0$, $X=(x,p)\in\mathbb{R}^{2n}$ and $\mathsf{V}(\theta,X)=\langle p,\mathcal{G}(\theta,x)\rangle$: the rank condition implies that derivatives of all orders of the smooth map $\theta\mapsto\mathsf{V}(\theta,X)$ never vanish at the same point, so that its zeroes are isolated and the set $\{\theta\in S^1\,,\ \mathsf{V}(\theta,X)=0\}$ is finite and *a fortiori* has measure zero; hence H is continuously differentiable outside $\{p=0\}$. Since $0\leq H(x,p)\leq k\|p\|$ for some local constant k the derivative of H^2 is zero at all points $(x,0)$ and, since (36) implies that the norm of $\mathrm{d}H(x,p)$ at neighboring points where $p\neq 0$ is bounded, the derivative of H^2 at these points tends to zero as $p\to 0$. H^2 is therefore continuously differentiable everywhere. ◻

The map H fails in general to be twice differentiable on $\widetilde{\mathcal{Z}}$; indeed its first derivative fails to be Lipschitz continuous, but we have the following estimate of its modulus of continuity, in the less degenerated case.

Theorem 3.15. *Assume that the rank condition (30) holds and that*

(i) *for $(x,p)\in\mathrm{T}^*M$, $p\neq 0$, there is at most one $\theta\in S^1$ such that $\langle p,\mathcal{G}(\theta,x)\rangle=0$, and $\langle p,\dfrac{\partial\mathcal{G}}{\partial\theta}(\theta,x)\rangle$ does not vanish at the same point,*

(ii) *for all $(\theta,x)\in S^1\times M$, one has $\operatorname{rank}\mathcal{G}(\theta,x)=m$,*

then any point $(\bar{x},\bar{p})$ has a constant c and a coordinate neighborhood in T^*M such that for all X and Y in $\mathbb{R}^{2n}$, coordinates of points in the neighborhood,

$$\|\mathrm{d}H(Y)-\mathrm{d}H(X)\|\leq c\,\|X-Y\|\,\ln\frac{1}{\|X-Y\|}\ . \tag{38}$$

In the left-hand side of (38), and (40) below, $\|.\|$ stands for the operator norm in coordinates, see §2.3. As for Theorem 3.13, we first prove a more generic result: Proposition 3.16 below, whose notations are these of Proposition 3.14, independent from the rest of the paper; its proof is in the Appendix to the paper.

Proposition 3.16. *Let V satisfy the following assumptions:*

$$\left.\begin{array}{ll}\text{(a)} & \mathsf{V}(\theta,X)=0 \text{ for all } (\theta,X)\in S^1\times O^{\mathsf{d}} \text{ such that } \mathsf{V}(\theta,X)=0,\\ \text{(b)} & \text{for any } X\in O^{\mathsf{d}}, \text{ there is at most one } \theta \text{ such that } \mathsf{V}(\theta,X)=0,\\ \text{(c)} & \mathsf{V} \text{ and } \partial\mathsf{V}/\partial\theta \text{ do not vanish simultaneously,}\end{array}\right\} \tag{39}$$

and let $\bar{X}$ be in $\widetilde{\mathcal{Z}}$. There is a neighborhood U of $\bar{X}$ in O^{d}, and a constant $K>0$ such that, for all X,Y in U,

$$\Big\|\mathrm{d}\mathsf{H}(X)-\mathrm{d}\mathsf{H}(Y)\Big\|\leq K\,\|X-Y\|\,\ln\frac{1}{\|X-Y\|}\ . \tag{40}$$

Proof of Theorem 3.15. Smoothness outside $\widetilde{\mathcal{Z}}$ is obvious from the expression (5) of H; inequality (38) is a consequence of Proposition 3.16, applied with $\mathsf{d}=2n$, $X=(x,p)\in\mathbb{R}^{2n}$, $\mathsf{V}(\theta,X)=\langle p,\mathcal{G}(\theta,x)\rangle$ and O^{d} a neighborhood of a point of $\widetilde{\mathcal{Z}}\setminus\{p=0\}$; it is clear that points (i) and (ii) imply the three conditions (39). ◻

Remark 3.17 (Finsler geometry). If H^2 was an least *twice* continuously differentiable, with a positive definite Hessian with respect to p, so would be N^2 (see (32)), and it would define a (reversible) *Finsler metric* [3] on M. The lack of differentiability calls for further developments.

3.5. Application to the minimum time problem. Fix two points x_0, x_1 in M and consider the time optimal problem associated to (1) for $\varepsilon > 0$:

$$(\mathcal{P}_\varepsilon),\ \varepsilon > 0: \qquad \left.\begin{array}{l} \dot{x}(t) = \mathcal{G}(t/\varepsilon, x(t))u(t),\ u(t) \in B^m,\ t \in [0,T], \\ x(0) = x_0,\ x(T) = x_1 \end{array}\right\} \ \min T\,, \tag{41}$$

and the time optimal problem associated to the average system:

$$(\mathcal{P}_0): \qquad \left.\begin{array}{l} \dot{x}(t) \in \mathcal{E}(x(t)),\ t \in [0,T], \\ x(0) = x_0,\ x(T) = x_1 \end{array}\right\} \ \min T\,. \tag{42}$$

Call $T_\varepsilon(x_0, x_1)$ the minimum time for $(\mathcal{P}_\varepsilon)$, $\varepsilon > 0$ and $T_0(x_0, x_1)$ the one for $(\mathcal{P}_0)$; when no confusion arises, we write T_ε and T_0.

Let us develop (41)–(42): concerning (41), T_ε is the infimum of the set of T's such that there is an admissible control $u(.)\colon [0,T] \to B^m$, and $x(.)\colon [0,T] \to M$ satisfying $x(0) = x_0$, $x(T) = x_1$ and $\dot{x}(t) = \mathcal{G}(t/\varepsilon, x(t))u(t)$ for almost all t; Proposition 3.11, point 2 implies that this set is nonempty, hence T_ε is finite. Concerning (42), T_0 is the infimum of the set of T's such that there is $x(.)\colon [0,T] \to M$ satisfying $x(0) = x_0$, $x(T) = x_1$ and $\dot{x}(t) \in \mathcal{E}(x(t))$ for almost all t, T_0 is finite from Proposition 3.12, point 2. A *solution* to $(\mathcal{P}_\varepsilon)$ (resp. to $(\mathcal{P}_0)$) is $x(.), u(.)$ (resp. $x(.)$) as above with $T = T_\varepsilon$ (resp. $T = T_0$). In general, the minimum T_ε or T_0 need not be reached, i.e. there need not be a solution.

LEMMA 3.18. *Assume the rank condition (30).*

1. There is a neighborhood $\mathcal{W}$ of any x_1 and two constants $\alpha_0 > 0$ and $C_3 > 0$ such that, for all y in $\mathcal{W}$, $T_\varepsilon(y, x_1) \leq 2\pi\varepsilon + C_3 d(x_1, y)$.

2. For any x_0, x_1', x_1 in M, one has $T_\varepsilon(x_0, x_1) \leq T_\varepsilon(x_0, x_1') + T_\varepsilon(x_1', x_1) + 2\pi\varepsilon$.

Proof. Apply Proposition 3.11, point 1 with $T = 2\pi$, using as a distance in $\mathcal{W}$ the Euclidean norm in some coordinates: for any two points y, y' in $\mathcal{W}$ such that $\|y - y'\| \leq \alpha_0$, there is a control defined on $[0, 2\pi]$, with L^∞ norm smaller than $c_3\|y - y'\|$ that brings y' at time 0 to y at time 2π for system (31); rescaling time and control by ε yields, if $c_3\|y - y'\| \leq \varepsilon$, a control with L^∞ norm less than 1 that brings y' at time 0 to y at time $2\pi\varepsilon$ for system (1) and hence, by concatenating controls and using periodicity of $\mathcal{G}$, for any positive integer k, a control with L^∞ norm less than 1 that brings y' at time 0 to y at time $2k\pi\varepsilon$ for system (1) if $c_3\|y - y'\| \leq k\varepsilon$. In other words, $T_\varepsilon(y', y) \leq 2\pi(\varepsilon + c_3\|y - y'\|)$. Take $y' = x_1$ and $2\pi c_3/C_3$ the ratio between the Euclidean norm and the distance d; this proves point 1. Point 2 follows from using periodicity of $\mathcal{G}$ and concatenating controls while inserting a zero control between time $T_\varepsilon(y', y)$ and the next multiple of 2π. ∎

THEOREM 3.19 (limit of minimum time). *Assume the rank condition (30).*

1. *T_ε is bounded as $\varepsilon \to 0$ and* $\limsup_{\varepsilon\to 0} T_\varepsilon \leq T_0$.

2. *If, for $\varepsilon > 0$ small enough, each $(\mathcal{P}_\varepsilon)$ has a solution $x_\varepsilon : [0, T_\varepsilon] \to M$ and there exists a compact $\mathbb{K} \subset M$ such that $x_\varepsilon([0, T_\varepsilon]) \subset \mathbb{K}$ for all $\varepsilon > 0$ small enough, then all accumulation points of the compact family $(x_\varepsilon(.))_{\varepsilon>0}$ in $C^0([0, T_0], M)$ are solutions of $(\mathcal{P}_0)$ and* $\lim_{\varepsilon\to 0} T_\varepsilon = T_0$.

Proof. Consider a minimizing sequence for problem $(\mathcal{P}_0)$, i.e. solutions x^k : $[0, T_0 + \beta_k] \to M$ of the average system (4) with (β_k) a sequence of positive numbers

that tends to zero and $x^k(0) = x_0$, $x^k(T_0 + \beta_k) = x_1$ for all k. For each x^k, there is, according to Theorem 3.7, a family $(\, x^k_\varepsilon(.)\,)_{\varepsilon>0}$ such that each $x^k_\varepsilon(.)$ is a solution of (1) with $x^k_\varepsilon(0) = x_0$ and $d(x^k_\varepsilon(t), x^k(t)) \leq c_1\varepsilon$ for all t in $[0, T_0 + \beta_k]$. In particular $d(x^k_\varepsilon(T_0 + \beta_k), x^k(t)_1) \leq c_1\varepsilon$. Now, from Lemma 3.18 stated below, $T_\varepsilon(\, x^k_\varepsilon(T_0 + \beta_k)\,, x_1) \leq (2\pi + c_1C_3)\varepsilon$; hence, from the second point of that lemma (with $x'_1 = x^k_\varepsilon(T_0 + \beta_k)\,$), one has $T_\varepsilon = T_\varepsilon(x_0, x_1) \leq T_0 + \beta_k + (4\pi + c_1C_3)\varepsilon$ and, letting k go to infinity, $T_\varepsilon \leq T_0 + (4\pi + c_1C_3)\varepsilon$; this implies Point 1.

In Point 2, first extend x_ε on $[0, \overline{T}]$, with $\overline{T}$ an upperbound of T_ε, by taking $x_\varepsilon(t) = x_1$ for all $t \in [T_\varepsilon, \overline{T}]$; any sequence $(x_{\varepsilon_k}(.))_{k\in\mathbb{N}}$ with $\lim \varepsilon_k = 0$ is compact in $C^0([0, \overline{T}], M)$; take a convergent subsequence such that T_{ε_k} also converges to some T^*; from Theorem 3.7, the uniform limit is a solution of the average system (4), and it goes through x_0 at time 0 and x_1 at time T^*, hence $T^* \geq T_0$ by definition of T_0. Since T^* can be any accumulation point of (T_ε) as $\varepsilon \to 0$, this, together with Point 1 ends the proof of Point 2. ◻

Let us now write the Pontryagin Maximum Principle (PMP) [22] both for $(\mathcal{P}_\varepsilon)$, $\varepsilon > 0$ and for $(\mathcal{P}_0)$ and see how they are related.

The *extremals* of problem $(\mathcal{P}_\varepsilon)$, $\varepsilon > 0$, are absolutely continuous maps $t \mapsto (x(t), p(t))$ solution to

$$\dot{p} = -\frac{\partial H_\varepsilon}{\partial x}\,, \quad \dot{x} = \frac{\partial H_\varepsilon}{\partial p} \qquad \text{with} \qquad H_\varepsilon(t, p, x) = \|\langle p, \mathcal{G}(t/\varepsilon, x)\rangle\|\,, \tag{43}$$

whose right-hand side is discontinuous on $\mathcal{S}_\varepsilon = \{(x, p, t),\ \langle p, \mathcal{G}(t/\varepsilon, x)\rangle = 0\}$ (the "switching surface"), where it is in fact not defined.

The *extremals* of $(\mathcal{P}_0)$ are absolutely continuous $t \mapsto (x(t), p(t))$ solution to

$$\dot{p} = -\frac{\partial H}{\partial x}\,, \quad \dot{x} = \frac{\partial H}{\partial p}\ . \tag{44}$$

with H given by (5). The right-hand sides are continuous according to Theorem 3.13.

THEOREM 3.20. *If an absolutely continuous map $t \mapsto \overline{x}(t)$ defined on $[0, \overline{T}]$ is a solution of $(\mathcal{P}_\varepsilon)$, $\varepsilon > 0$, (resp. of $(\mathcal{P}_0)$), then there exists $t \mapsto \overline{p}(t)$ defined on $[0, \overline{T}]$ such that $t \mapsto (\overline{p}(t), \overline{x}(t))$ is an extremal of $(\mathcal{P}_\varepsilon)$, $\varepsilon > 0$ (resp. of $(\mathcal{P}_0)$).*

Proof. Problem $(\mathcal{P}_\varepsilon)$, $\varepsilon > 0$ deals with a classical smooth control system; according to [22, 1], the pseudo-Hamiltonian is $h(t, x, p, u) = \langle p, \mathcal{G}(t/\varepsilon, x)\, u\rangle$; an extremal is a curve on the co-tangent bundle solution, in local coordinates, of:

$$\begin{array}{rcl} \dot{p} & = & -\frac{\partial h}{\partial x}(t, x, p, u^*) \ = \ -\langle p, \frac{\partial \mathcal{G}}{\partial x} u^*\rangle, \\ \dot{x} & = & \frac{\partial h}{\partial p}(t, x, p, u^*) \ = \ \mathcal{G}\, u^*, \end{array} \tag{45}$$

with $u^*(t)$ a control that maximizes the pseudo-Hamiltonian for almost all time; it is defined by $u^* = \frac{\langle p, \mathcal{G}\rangle}{\|\langle p, \mathcal{G}\rangle\|}$ if $\langle p, \mathcal{G}(t/\varepsilon, x)\rangle \neq 0$; the maximized Hamiltonian $H_\varepsilon(t, p, x) = \max_u h(t, x, p, u)$ is the one in (43) and (45) is then the differential equation (43), whose right-hand side is discontinuous at points where $\langle p, \mathcal{G}(t/\varepsilon, x)\rangle$ vanishes.

Let us now turn to $(\mathcal{P}_0)$. Since the set of admissible velocities is not a priori smooth with respect to the state variable we use a non-smooth version of the Pontryagin maximum principle for differential inclusions, that we recall for self-containedness:

> Theorem 9.1 in [11, Chapter 4]: *if $\dot{x} \in \mathcal{E}(x)$ is a locally Lipschitz differential inclusion and $t \mapsto \overline{x}(t)$ is an absolutely continuous function defined on $[0, \overline{T}]$ solution to the problem (42), then there exists $t \mapsto \overline{p}(t)$ defined on $[0, \overline{T}]$ such that $(-\dot{\overline{p}}, \dot{\overline{x}}) \in \partial_C H(\overline{x}, \overline{p})$ for almost all $t \in [0, \overline{T}]$ with $H(x, p) = \max_{v\in\mathcal{E}(x)}\langle p, v\rangle$ and $\partial_C H$ the generalized gradient of H.*

The set-valued map $\mathcal{E}(.)$ in (3) is indeed locally Lipschitz: in local coordinates , for x_1, x_2 in $\mathbb{R}^n$, denoting by δ the Hausdorff distance between two sets, one has:

$$\begin{aligned}\delta\left(\mathcal{E}(x_1), \mathcal{E}(x_2)\right) &= \max\left\{\sup_{v_1\in\mathcal{E}(x_1)} \inf_{v_2\in\mathcal{E}(x_2)} \|v_1 - v_2\|, \sup_{v_2\in\mathcal{E}(x_2)} \inf_{v_1\in\mathcal{E}(x_1)} \|v_1 - v_2\|\right\} \\ &= \max_{\|\mathcal{U}_1\|_\infty\le 1} \min_{\|\mathcal{U}_2\|_\infty\le 1} \|\overline{\mathcal{G}}(x_1, \mathcal{U}_1) - \overline{\mathcal{G}}(x_2, \mathcal{U}_2)\| \;\le\; \mathrm{Lip}\,\mathcal{G}\;\|x_1 - x_2\|.\end{aligned}$$

According to (8), the Hamiltonian H defined in the above quoted theorem coincides with the map H defined in (5). ☐

Remark 3.21. This result and Theorem 3.19 have two interpretations:
1. They prove that the operations of *averaging* and *computing the Hamiltonian for the minimum time problem* commute. Indeed, the Hamiltonian H was obtained by applying the maximum principle to problem (42), i.e. minimum time for the average system (4), but it also the average of the one in (43) with respect to the fast variable.
2. They prove indirectly an averaging result for the minimum time control problem (42); the averaging techniques in [10] do not apply to minimum time for they require smoothness of the Hamiltonian, while averaging is used in [14, 13] for minimum time with only partial theoretical justifications but numerical evidence of efficiency.

Let us now focus on the differential equations (44) that govern the extremals of $(\mathcal{P}_0)$. It is of great importance to know whether it defines a Hamiltonian flow on T^*M, *i.e.* whether solutions trough all initial conditions are unique or not. Its right-hand side is continuous because, from Theorem 3.13, H is continuously differentiable; this ensures existence of solutions. We saw that H is smooth (C^∞) on $\mathrm{T}^*M \setminus \widetilde{\mathcal{Z}}$ (see (33)), hence solutions through points outside $\widetilde{\mathcal{Z}}$ are always unique. The following result gives uniqueness of solutions even on $\widetilde{\mathcal{Z}}$ in the less degenerated case possible.

THEOREM 3.22 (Hamiltonian flow for $(\mathcal{P}_0)$). *Assume that the rank condition (30) holds, as well as conditions (i) and (ii) in Theorem 3.15. Then the differential equation* (44) *has a unique solution from any initial condition.*

Proof. For an autonomous ODE $\dot z = f(z)$ in a finite dimensional space, where f satisfies $\|f(z_1) - f(z_2)\| \le \omega(\|z_1 - z_2\|)$ with $\omega : [0, +\infty) \to [0, +\infty)$ non-decreasing, Kamke uniqueness Theorem [15, chap. III, Th. 6.1] states that uniqueness of solutions holds if $\int_0^\alpha \frac{\mathrm{d}u}{\omega(u)} = +\infty$ for arbitrarily small $\alpha > 0$. From Theorem 3.15, we are in this case with $\omega(u) = c\,u \ln(1/u)$, and $\int \frac{\mathrm{d}u}{\omega(u)} = -c \ln\ln\frac{1}{u}$. ☐

The sufficient condition given by this theorem is certainly not the best possible, but turns out to be applicable to the control of orbit transfer with low thrust, see §5. Point (ii) is very mild and only states that the control vector fields are linearly independent. Point (i) is more artificial: the fact that $\langle p, \mathcal{G}(\theta, x)\rangle = 0$ has at most one solution θ has to be checked by hand, while the fact that $\partial\mathcal{G}/\partial\theta$ does not vanish at the same time, which is true for the Kepler problem and used in [9] to show that the switchings in (43) are always "π-singularities", i.e. the control u^* switches to its opposite, is equivalent to the rank condition

$$\mathrm{rank}\left\{\mathcal{G}(\theta, x), \frac{\partial\mathcal{G}}{\partial\theta}(\theta, x)\right\} = \dim\Big(\mathrm{Range}\,\mathcal{G}(\theta, x) + \mathrm{Range}\,\frac{\partial\mathcal{G}}{\partial\theta}(\theta, x)\Big) = n\ . \qquad (46)$$

Proving existence of a flow for (44) in more general situations is obviously an interesting program to be pursued.

4. Kepler control systems. We call *Kepler control system* with small control a family of control system on $S^1 \times M$ of the form

$$(\mathcal{K}_\varepsilon) \quad \left\{ \begin{array}{rcl} \dot\theta &=& \omega(\theta,x) + g(\theta,x)\,v \\ \dot x &=& G(\theta,x)\,v \end{array} \right. \quad , \quad \|v\| \le \varepsilon \ , \tag{47}$$

where G and g can be viewed, with the same convention is in (1), as $n \times m$ and $1 \times m$ matrices smoothly depending on (θ, x) and ω is a smooth function $S^1 \times M \to \mathbb{R}$ that remains larger than a strictly positive constant:

$$\omega(\theta,x) \ge k_\omega > 0 \quad \forall (\theta,x) \in S^1 \times M \ . \tag{48}$$

In fact, this is an affine control system on $S^1 \times M$

$$\dot\xi = f_0(\xi) + \sum_{i=1}^m u_i f_i(\xi) \tag{49}$$

with $\xi = (\theta, x)$, $f_0 = \omega \frac{\partial}{\partial\theta}$ and, for $1 \le i \le m$, the smooth vector field f_i is represented by the i^{th} column of the matrix notations G and g.

In (49), if all solutions of $\dot\xi = f_0(\xi)$ are periodic but ξ is not a priori decomposed as a product, additional conditions are needed for the orbits to induce a nice foliation that splits the state manifold into a product $M \times S^1$.

4.1. Relation with fast oscillating systems. Let us define a new time λ in order to transform the first equation in (47) into a "fast oscillating system" of the type (1). For a solution $t \mapsto (\theta(t), x(t))$, let $\Theta(t)$ be the cumulated angle *i.e.* $\Theta(.)$ is continuous $[0,T] \to \mathbb{R}$ with $\Theta(t) \equiv \theta(t) \bmod 2\pi$ for all t and $\Theta(0) \in [0, 2\pi)$, and let

$$\lambda = \mathcal{R}(t) \stackrel{\Delta}{=} \varepsilon\,(\Theta(t) - \Theta(0)) \ . \tag{50}$$

Taking ε_0 small enough that $|\omega(\theta,x) + \varepsilon\,g(\theta,x)\,u| > k_\omega/2$ for $x \in \mathbb{K}$, $\|u\| \le 1$, $\varepsilon < \varepsilon_0$, one has $\mathrm{d}\mathcal{R}/\mathrm{d}t > \varepsilon\,k_\omega/2$ hence $\mathcal{R}$ is strictly increasing and one-to-one, and

$$\frac{k_\omega}{2}\varepsilon\,t \le \mathcal{R}(t) \le k_\omega\,\varepsilon\,t \quad \text{with } k_\omega = \sup_{S^1\times\mathbb{K}} \omega + \varepsilon_0 \sup_{S^1\times\mathbb{K}} \|g\| \,. \tag{51}$$

Define, from $v(.)$, the control $\lambda \mapsto \widehat u(\lambda) = v(\mathcal{R}^{-1}(\lambda))/\varepsilon$; then $\widetilde x(\lambda) = x(\mathcal{R}^{-1}(\lambda))$ satisfies

$$(\widetilde\Sigma_{\theta_0,\varepsilon}) \qquad \frac{\mathrm{d}\widetilde x}{\mathrm{d}\lambda} = \frac{G(\theta_0 + \frac{\lambda}{\varepsilon}, \widetilde x)\,\widehat u}{\omega(\theta_0 + \frac{\lambda}{\varepsilon}, \widetilde x) \ + \ \varepsilon\,g(\theta_0 + \frac{\lambda}{\varepsilon}, \widetilde x)\,\widehat u} \ , \quad \|\widehat u\| \le 1 \ . \tag{52}$$

Except for the term $\varepsilon g \widehat u$ in the denominator, this is a fast oscillating system (1) with a suitable definition of $\mathcal{G}$. We will now apply the concepts and results of §3.

4.2. Average control system. The definition uses ϖ defined by

$$\frac{1}{\varpi(x)} = \frac{1}{2\pi}\int_0^{2\pi} \frac{\mathrm{d}\theta}{\omega(\theta,x)} \ . \tag{53}$$

DEFINITION 4.1 (Average control system of Kepler control systems). *The average control system of the Kepler control system (47) is the differential inclusion*

$$\dot x \in \mathcal{E}(x) \tag{54}$$

with $\mathcal{E}$ *defined by (3) using* $\overline{\mathcal{G}} : M \times L^\infty(S^1, \mathbb{R}^m) \to \mathrm{T}M$ *defined by*

$$\overline{\mathcal{G}}(x, \mathcal{U}) \;=\; \overline{\omega}(x)\, \frac{1}{2\pi}\!\int_0^{2\pi} \frac{G(\theta, x)}{\omega(\theta, x)}\, \mathcal{U}(\theta)\mathrm{d}\theta \tag{55}$$

instead of (2). *Solutions are defined as in Definition 3.2.*

Remark 4.2. This Definition 3.2 applied to (52), which is equivalent to (47) via time changes, *except*:

(i) the term $\varepsilon g \widehat{u}$ in the denominator of (52) has been discarded,

(ii) the right-hand side has been multiplied by $\overline{\omega}(x)$ so that the "time" of the average system is equivalent to ε times the "time" of the control system (47) as $\varepsilon \to 0$.

4.3. Convergence Theorem. The counterpart of Theorem 3.7 is:

THEOREM 4.3 (Convergence for Kepler control systems).

1. *Let* $x_0(.) : [0, T] \to M$ *be en arbitrary solution of* (54) *and* $\theta^0 \in S^1$. *There exist a family of measurable functions* $\overline{u}_\varepsilon(.) : [0, T] \to B^m$, *indexed by* $\varepsilon > 0$, *and positive constants* c, ε_0, *such that, if* $t \mapsto (\theta_\varepsilon(t), x_\varepsilon(t))$ *is the solution of* (47) *with control* $u = \overline{u}_\varepsilon(t)$ *and initial condition* $(\theta_\varepsilon(0), x_\varepsilon(0)) = (\theta^0, x_0(0))$, *it is defined on* $[0, T/\varepsilon]$ *for* ε *smaller than* ε_0 *and*

$$d(\, x_\varepsilon(t)\, ,\, x_0(\varepsilon t)\,) < c\, \varepsilon\,, \quad t \in [0, \frac{T}{\varepsilon}]\,, \quad 0 < \varepsilon < \varepsilon_0\,, \tag{56}$$

thus $\tau \mapsto x_\varepsilon(\tau/\varepsilon)$ *converges uniformly on* $[0, T]$ *to* $\tau \mapsto x_0(\tau)$ *when* ε *tends to zero.*

2. *Let* $\mathbb{K}$ *be a compact subset of* M, $(\varepsilon_n)_{n\in\mathbb{N}}$ *a decreasing sequence of positive real numbers converging to zero, and* $\big(\theta_n(.), x_n(.)\big) : [0, T/\varepsilon_n] \to S^1 \times \mathbb{K}$ *a solution of system (47) for each* n, *with* $\varepsilon = \varepsilon_n$ *and some control* $u = u_n(t)$, $u_n(.) \in L^\infty([0, T/\varepsilon_n], \mathbb{R}^m)$, $\|u_n(.)\|_\infty \leq 1$. *Then the sequence* $\big(\tau \mapsto (x_n(\tau/\varepsilon_n))\big)_{n\in\mathbb{N}}$ *is compact for the topology of uniform convergence on* $[0, T]$ *and the limit of any converging sub-sequence is a solution* $x^*(.)$ *of the average differential inclusion* (54).

Proof. This proof is done as if M was $\mathbb{R}^n$, d the Euclidean distance and all vector fields had a common compact support, hence all maps a global Lipschitz constant and a global bound; it is left to the reader to check, as in the proof of Theorem 3.7, that it translates on M with any distance d described in §2.3.

Since $(\widetilde{\Sigma}_{\theta_0,\varepsilon})$ in (52) can be re-written as

$$\frac{\mathrm{d}x}{\mathrm{d}\lambda} = \Big(1 - \frac{\varepsilon\, g(\theta_0 + \frac{\lambda}{\varepsilon}, x)\, u}{\omega(\theta_0 + \frac{\lambda}{\varepsilon}, x) + \varepsilon\, g(\theta_0 + \frac{\lambda}{\varepsilon}, x)\, u}\Big) \frac{G(\theta_0 + \frac{\lambda}{\varepsilon}, x)\, u}{\omega(\theta_0 + \frac{\lambda}{\varepsilon}, x)}\,, \tag{57}$$

the norm of the difference between the right-hand sides of $(\widetilde{\Sigma}_{\theta_0,\varepsilon})$ and of

$$(\widehat{\Sigma}_{\theta_0,\varepsilon}) \qquad \frac{\mathrm{d}\widehat{x}}{\mathrm{d}\lambda} = \frac{G(\theta_0 + \frac{\lambda}{\varepsilon}, \widehat{x})\, \widehat{u}}{\omega(\theta_0 + \frac{\lambda}{\varepsilon}, \widehat{x})}\;, \quad \|\widehat{u}\| \leq 1 \tag{58}$$

is bounded by $k\,\varepsilon$ for some $k > 0$ depending on the bound and Lipschitz constant of G and ω; classical theorems on smooth dependence of solutions on the equation yield some c' such that, if two solutions $\widetilde{x}_\varepsilon(.)$ of $(\widetilde{\Sigma}_{\theta_0,\varepsilon})$ and $\widehat{x}_\varepsilon(.)$ of $(\widehat{\Sigma}_{\theta_0,\varepsilon})$ are defined on the interval $[0, \Lambda]$, with the same control $\lambda \mapsto u(\lambda)$ and $\widetilde{x}_\varepsilon(0) = \widehat{x}_\varepsilon(0)$, then

$$d(\widetilde{x}_\varepsilon(\lambda), \widehat{x}_\varepsilon(\lambda)) \leq c'\, \varepsilon \;\text{ for all } \lambda \in [0, \Lambda]\,. \tag{59}$$

Let $\tau \mapsto x_0(\tau)$ be a solution of (54) on $[0,T]$. Then $\lambda \mapsto \widehat{x}_0(\lambda)$, with

$$\lambda = \mathcal{P}(\tau) = \int_0^\tau \overline{\varpi}\,(x_0(\mathfrak{t}))\,\mathrm{d}\mathfrak{t} \tag{60}$$

and $x_0(\tau) = \widehat{x}_0(\mathcal{P}(\tau))$ for all τ, is a solution on $[0, \mathcal{P}(T)]$ of

$$\frac{\mathrm{d}\widehat{x}_0}{\mathrm{d}\lambda} \in \frac{1}{\overline{\varpi}(\widehat{x}_0)}\mathcal{E}(\widehat{x}_0)\ . \tag{61}$$

This is, according to (55), the average system of the fast oscillating (58); Theorem 3.7 (Point 1) yields a family of controls $\widehat{u}_\varepsilon$ such that the solutions $\widehat{x}_\varepsilon(.)$ of $(\widehat{\Sigma}_{\theta_0,\varepsilon})$ with initial condition $\widehat{x}_0(0)$ and control $\widehat{u}_\varepsilon$ converge to $\widehat{x}_0(.)$ uniformly on $[0,\mathcal{P}(T)]$ with a distance less that $c''\varepsilon$. For each ε, the solution $\widetilde{x}_\varepsilon(.)$ of $(\widetilde{\Sigma}_{\theta_0,\varepsilon})$ with same initial condition and control satisfies (59) with $\Lambda = \mathcal{P}(T)$; then define

$$t = \mathfrak{T}(\lambda) = \frac{1}{\varepsilon}\int_0^\lambda \frac{\mathrm{d}\ell}{\omega(\theta_0+\frac{\ell}{\varepsilon},\widetilde{x}_\varepsilon(\ell)) + \varepsilon\, g(\theta_0+\frac{\ell}{\varepsilon},\widetilde{x}_\varepsilon(\ell))\,\widehat{u}_\varepsilon(\ell)} \tag{62}$$

and the controls $t \mapsto \overline{u}_\varepsilon(t)$ by $\widehat{u}_\varepsilon(\lambda) = \overline{u}_\varepsilon(\mathfrak{T}(\lambda))$; the solutions $x_\varepsilon(.)$ of (47) with these controls are given by $\widetilde{x}_\varepsilon(\lambda) = x_\varepsilon(\mathfrak{T}(\lambda))$, and one therefore has

$$d(x_\varepsilon(\mathfrak{T}\circ\mathcal{P}(\tau)), x_0(\tau)) < (c'+c'')\varepsilon, \quad \tau \in [0,T]\,. \tag{63}$$

Let us now study $\mathfrak{T}\circ\mathcal{P}(\tau)$. From (60) and (62), one has

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\tau}\mathfrak{T}\circ\mathcal{P}(\tau) &= \frac{1}{\varepsilon}\,\frac{\overline{\varpi}(x_0(\tau))}{\omega(\theta_0+\frac{\lambda}{\varepsilon},\widetilde{x}_\varepsilon(\lambda)) + \varepsilon\, g(\theta_0+\frac{\lambda}{\varepsilon},\widetilde{x}_\varepsilon(\lambda))\,\widehat{u}_\varepsilon(\lambda)} \\
&= \frac{1}{\varepsilon}\left(\frac{\overline{\varpi}(x_0(\tau))}{\omega(\theta_0+\frac{\lambda}{\varepsilon},\widehat{x}_0(\lambda))} + \rho(\varepsilon,\theta_0,\tau)\right) \quad \text{with} \quad \lambda = \mathcal{P}(\tau)\,,
\end{aligned}$$

where ρ contains a term for the difference between this fraction with or without $\varepsilon g\widehat{u}_\varepsilon$ in the denominator, of order ε by a computation similar to (57), and a term for the difference between $\widetilde{x}(\lambda)$ and $\widehat{x}(\lambda)$ as an argument of ω, of order ε by (59) and Lipschitz continuity of ω; hence $|\rho|/\varepsilon$ is bounded by a constant independent of ε. Finally,

$$\int_0^\tau \frac{\overline{\varpi}(x_0(\mathfrak{t}))\,\mathrm{d}\mathfrak{t}}{\omega(\theta_0+\frac{\mathcal{P}(\mathfrak{t})}{\varepsilon},\widehat{x}_0(\mathcal{P}(\mathfrak{t})))} = \int_0^{\mathcal{P}(\tau)} \frac{\mathrm{d}\lambda}{\omega(\theta_0+\frac{\lambda}{\varepsilon},\widehat{x}_0(\lambda))} = \frac{1}{2\pi}\iint_{\substack{\theta\in S^1\\ 0\le\lambda\le\mathcal{P}(\tau)}} \frac{\mathrm{d}\lambda\,\mathrm{d}\theta}{\omega(\theta_0+\frac{\lambda}{\varepsilon},\widehat{x}_0(\lambda))}\,,$$

and, since $\overline{\varpi}(x_0(.)) = \overline{\varpi}(\widehat{x}_0(\mathcal{P}(.)))$, and with $\mu(\theta)$ is defined in §2.4,

$$\begin{aligned}
\tau = \int_0^\tau \frac{\overline{\varpi}(x_0(\mathfrak{t}))\,\mathrm{d}\mathfrak{t}}{\overline{\varpi}(\widehat{x}_0(\mathcal{P}(\mathfrak{t})))} = \int_0^{\mathcal{P}(\tau)} \frac{\mathrm{d}\lambda}{\overline{\varpi}(\widehat{x}_0(\lambda))} &= \frac{1}{2\pi}\iint_{\substack{\theta\in S^1\\ 0\le\lambda\le\mathcal{P}(\tau)}} \frac{\mathrm{d}\lambda\,\mathrm{d}\theta}{\omega(\theta_0+\frac{\lambda}{\varepsilon}+\theta,\widehat{x}_0(\lambda))} \\
&= \frac{1}{2\pi}\iint_{\substack{\theta\in S^1\\ \varepsilon\mu(\theta)\le\ell\le\mathcal{P}(\tau)+\varepsilon\mu(\theta)}} \frac{\mathrm{d}\ell\,\mathrm{d}\theta}{\omega(\theta_0+\frac{\ell}{\varepsilon},\widehat{x}_0(\ell-\varepsilon\mu(\theta)))}
\end{aligned}$$

hence the difference is less than $k\varepsilon$ for some k independent of ε; the above implies that $\left|\mathfrak{T}\circ\mathcal{P}(\tau) - \frac{\tau}{\varepsilon}\right|$ is bounded by a constant. Finally, since $x_\varepsilon(.)$ is Lipschitz continuous

with constant $2\varepsilon\sup\|G\|/k_\omega$, one has $d(x_\varepsilon(\mathfrak{T}\circ\mathcal{P}(\tau)),x_\varepsilon(\frac{\tau}{\varepsilon}))<c'''\varepsilon$ for some ε and finally (63) implies point 1 of the theorem, with $c=c'+c''+c'''$ in (62).

For point 2, consider $\big(\theta_n(.),x_n(.)\big):[0,T/\varepsilon_n]\to S^1\times\mathbb{K}$ a solution of system (47) with $\varepsilon=\varepsilon_n$ and some control $u=u_n(t)$. One associates to these, following (50)–(52) and setting $\lambda=\mathcal{R}_n(t)$ (we write $\mathcal{R}_n$ because $\mathcal{R}$ in (50) is constructed for system $(\widetilde{\Sigma}_{\theta_0,\varepsilon_n})$ and thus depends on n), solutions and control $\lambda\mapsto\widetilde{x}_n(\lambda)$ and $\lambda\mapsto\widetilde{u}_n(\lambda)$ of $(\widetilde{\Sigma}_{\theta_0,\varepsilon_n})$. The solutions $\lambda\mapsto\widehat{x}_n(\lambda)$ of $(\widehat{\Sigma}_{\theta_0,\varepsilon_n})$ with same control and initial condition satisfy, according to (59), $d(\widehat{x}_n(\lambda),\widetilde{x}_n(\lambda))<c'\varepsilon_n$. By Theorem 3.7 (Point 2), the sequence $(\widehat{x}_n)$ is compact and subsequences converge to solutions $\lambda\mapsto\widehat{x}_0(\lambda)$ of (61), hence the same subsequences of $(\widetilde{x}_n)$ converge as well, and, with $\tau=\mathcal{Q}(\lambda)\triangleq\int_0^\lambda\frac{\mathrm{d}\ell}{\overline{\omega}(\widehat{x}(\ell))}$, the maps $\tau\mapsto\widetilde{x}_n(\mathcal{Q}^{-1}(\tau))=x_n((\mathcal{Q}\circ\mathcal{R}_n)^{-1}(\tau))$ converge to a solution $\tau\mapsto x_0(\tau)=\widehat{x}_0(\mathcal{Q}^{-1}(\tau))$ of the average system (54), with distance less than $c''\varepsilon_n$ for some c''. Using the same argument as in Point 1 for $\mathfrak{T}\circ\mathcal{P}(\tau)$, one gets a bound for $|(\mathcal{Q}\circ\mathcal{R}_n)^{-1}(\tau)-\frac{\tau}{\varepsilon_n}|$ and, for some c''', $d(\,x_n((\mathcal{Q}\circ\mathcal{R}_n)^{-1}(\tau))\,,\,x_n(\frac{\tau}{\varepsilon}))\le c'''\varepsilon_n$. Point 2 is proved. ∎

4.4. Dimension of $\mathcal{E}(x)$. In §3.3, and in particular in Proposition 3.10, $\mathcal{G}$ can simply be replaced with G. It is however interesting to give a more intrinsic characterization of $r(\theta,x)$ and thus of $\dim\mathcal{E}(x)$.

Proposition 4.4. *If* (47) *and* (49) *represent the same control system, then*

$$\dim\Big(\sum_{j\in\mathbb{N}}\operatorname{Range}\frac{\partial^jG}{\partial\theta^j}(\theta,x)\Big)\\ =-1+\operatorname{rank}\Big(\{f_0(\theta,x)\}\cup\Big\{\operatorname{ad}^j_{f_0}f_k\,(\theta,x)\;,\;j\in\mathbb{N},1\le k\le m\Big\}\Big)\,.\qquad(64)$$

Proof. Straightforward computation using the fact that $f_0=\partial/\partial\theta$. ∎

Note that the right-hand side is $r(\theta,x)$. Proposition 3.10 applies, with this definition of r. In particular, the "full rank case" becomes:

Proposition 4.5. *If the vector fields f_0 and $\operatorname{ad}^j_{f_0}f_k$,r $1\le k\le m$, $j\in\mathbb{N}$ span the whole tangent space of $S^1\times M$, then $\mathcal{E}(x)$ has a nonempty interior for all x.*

4.5. The function $H(x,p)$. Instead of (5), H has to be taken as follows, with $\overline{\omega}$ defined in (53):

$$H(x,p)=\ \overline{\omega}(x)\ \frac{1}{2\pi}\int_0^{2\pi}\Big\|\Big\langle p,\frac{G(\theta,x)}{\omega(\theta,x)}\Big\rangle\Big\|\mathrm{d}\theta\;.\qquad(65)$$

The characterization of $\mathcal{E}(x)$ in Proposition 3.4 is unchanged. In the "full rank case", the results from §3.4 on the degree of differentiability apply without a change.

4.6. Application to the minimum time problem. As in §3.5, but for the Kepler system (47), let x_0, x_1 be fixed, call T_ε the minimum time such that, from *some* θ_0,θ_1, (θ_1,x_1) can be reached from (θ_0,x_0) in system $(\mathcal{K}_\varepsilon)$ (obviously $T_\varepsilon\to+\infty$ as $\varepsilon\to0$) and T_ε the minimum time such that x_1 can be reached from x_0 in the average system (54). The equivalent of Theorem 3.19, with a similar proof, using Theorem 4.3, is:

Theorem 4.6. *In the full rank case, one has* $\limsup_{\varepsilon\to0}\varepsilon T_\varepsilon\le T_0$ *(hence $\varepsilon\,T_\varepsilon$ is bounded as $\varepsilon\to0$). If, for all $\varepsilon>0$ small enough, there is a minimizing solution $(\theta_\varepsilon,x_\varepsilon):[0,T_\varepsilon]\to S^1\times M$ and they all remain in a common compact subset of M, then all accumulation points (as $\varepsilon\to0$) of the compact family $(\tau\to x_\varepsilon(\frac{\tau}{\varepsilon}))_{\varepsilon>0}$ in $C^0([0,T_0],M)$ are minimizing for the average system and* $\lim_{\varepsilon\to0}\varepsilon\,T_\varepsilon=T_0$.

The Hamiltonian for minimum time for the average system is given by (65); one has to perform the time scaling described in §4.1 to have a result like Theorem 3.22 and the simple "commutation between averaging and writing Hamiltonian" noted in Remark 3.21. Let us translate in terms of (47) the sufficient condition for existence of a Hamiltonian flow given by Theorem 3.22:

THEOREM 4.7. *In the full rank case, assume that* $\langle p, G(\theta, x)\rangle$ *and* $\langle p, \partial G/\partial\theta(\theta, x)\rangle$ *do not vanish simultaneously outside* $\{p = 0\}$*, that* $\theta \mapsto \langle p, G(\theta, x)\rangle$ *vanishes at most once for each* $(x, p) \in \mathrm{T}^*M$*,* $p \neq 0$*, and that* $\operatorname{rank} \mathcal{G}(\theta, x) = m$ *for each* $(\theta, x) \in S^1 \times M$*. Then* (44)*, with* H *given by* (65)*, has unique solutions through any initial condition.*

The discussion of Theorem 3.22 also applies to the above; let us mention that, once it has been checked that, for each (x, p), $\langle p, G(\theta, x)\rangle$ vanishes for at most one θ, the other conditions are guaranteed if, (46) holds with $\mathcal{G}$ replaced by G or, in terms of the vector fields in (49), if, for all $\xi = (\theta, x)$,

$$\operatorname{rank}\{f_0(\xi), f_1(\xi), \dots, f_m(\xi), \operatorname{ad}_{f_0} f_1(\xi), \dots, \operatorname{ad}_{f_0} f_m(\xi)\} = n + 1. \tag{66}$$

We prove in next section that the above conditions are true at least for the planar control 2-body problem.

5. Application to the controlled 2-body system. In this section we study some properties of the planar control system and demonstrate that it satisfies the condition of Theorem 3.22 on the domain of non-degenerated elliptic orbits.

5.1. Planar control 2-body system. It is classically described by some first integrals of the free movement —here the semi-major axis a and the eccentricity vector (e_x, e_y)— and one angle L following the dynamics; we restrict to the set of non-degenerated elliptic orbits rotating in the direct sense, i.e. the state space is $S^1 \times M$ with $M = \{(a, e_x, e_y) \in \mathbb{R}^3,\ a > 0 \text{ and } {e_x}^2 + {e_y}^2 < 1\}$. The control $u = (u_t, u_n)$ is expressed in the tangential-normal frame and the system reads:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} a \\ e_x \\ e_y \\ L \end{pmatrix} = \frac{1}{a^{3/2}}\begin{pmatrix} 0 \\ 0 \\ 0 \\ \mathtt{w}(e_x, e_y, L) \end{pmatrix} + \sqrt{a}\begin{pmatrix} 2\,a\,\mathtt{a}_a(e_x, e_y, L) & 0 \\ 2\,\mathtt{a}_x(e_x, e_y, L) & \mathtt{b}_x(e_x, e_y, L) \\ 2\,\mathtt{a}_y(e_x, e_y, L) & \mathtt{b}_y(e_x, e_y, L) \\ 0 & 0 \end{pmatrix}\begin{pmatrix} u_t \\ u_n \end{pmatrix} \tag{67}$$

with
$$\mathtt{w}(e_x, e_y, L) = \frac{(1 + e_x \cos L + e_y \sin L)^2}{(1 - e^2)^{3/2}},$$

$$\mathtt{a}_a(e_x, e_y, L) = \frac{\sqrt{1 + e^2 + 2\,e_x \cos L + 2\,e_y \sin L}}{\sqrt{1 - e^2}},$$

$$\mathtt{a}_x(e_x, e_y, L) = \frac{\sqrt{1 - e^2}}{\sqrt{1 + e^2 + 2\,e_x \cos L + 2\,e_y \sin L}}(e_x + \cos L),$$

$$\mathtt{a}_y(e_x, e_y, L) = \frac{\sqrt{1 - e^2}}{\sqrt{1 + e^2 + 2\,e_x \cos L + 2\,e_y \sin L}}(e_x + \cos L),$$

$$\mathtt{b}_x(e_x, e_y, L) = \frac{\sqrt{1 - e^2}}{\sqrt{1 + e^2 + 2\,e_x \cos L + 2\,e_y \sin L}} \times \frac{-2\,e_y + (e_x^2 - e_y^2 - 1)\sin L - 2\,e_x e_y \cos L}{1 + 2\,e_x \cos L + 2\,e_y \sin L},$$

$$\mathtt{b}_y(e_x,e_y,L) = \frac{\sqrt{1-e^2}}{\sqrt{1+e^2+2\,e_x\cos L+2\,e_y\sin L}} \times \frac{2\,e_x+(e_x^2-e_y^2+1)\cos L+2\,e_xe_y\sin L}{1+2\,e_x\cos L+2\,e_y\sin L}\,.$$

The eccentricity e is the norm of the eccentricity vector, $e=\sqrt{e_x^2+e_y^2}$. Low thrust translates into $\|u\|\leq\varepsilon$ for a small ε.

Remark 5.1. This is indeed a "Kepler control system" of the type (47) except that $\omega=\mathtt{w}/a^{3/2}$ is, although strictly positive, not bounded from below by a positive constant on $S^1\times M$. There is such a lowerbound if one replaces M by $M^{\bar c}=\{(a,e_x,e_y)\in\mathbb{R}^3,\ a>0 \text{ and } {e_x}^2+{e_y}^2<\bar c\}$ with $\bar c<1$. Strictly speaking, the results of the paper have to be applied in $M^{\bar c}$, $\bar c<1$. However, Theorems 4.3 or 4.7, for instance, may be applied in M because each statement may ultimately be restricted to a compact subset of M, itself included in some $M^{\bar c}$, $\bar c<1$.

The Hamiltonian that both defines the average system according to (6) and yields the Hamiltonian system governing extremals for minimum time is given by (65). Since $\int_0^{2\pi}\mathrm{d}L/\mathtt{w}(e_x,e_y,L)=2\pi$, it can be expressed as $H(a,e_x,e_y,p_a,p_{e_x},p_{e_y})=\sqrt{a}\mathcal{H}(e_x,e_y,ap_a,p_{e_x},p_{e_y})$ with

$$\mathcal{H}(e_x,e_y,A,X,Y) = \frac{1}{2\pi}\int_0^{2\pi}\|(A\,X\,Y)\mathtt{G}(e_x,e_y,L)\|,$$

$$\mathtt{G}(e_x,e_y,L) = \begin{pmatrix} 2\,\mathtt{a}_a/\mathtt{w} & 0\\ 2\,\mathtt{a}_x/\mathtt{w} & \mathtt{b}_x/\mathtt{w}\\ 2\,\mathtt{a}_y/\mathtt{w} & \mathtt{b}_y/\mathtt{w}\end{pmatrix}.$$

5.2. Hamiltonian flow. Theorem 4.7 applies to this system. Indeed:

PROPOSITION 5.2. *Fore each (e_x,e_y,a) with ${e_x}^2+{e_y}^2<1$ and $a>0$, and each $(A,X,Y)\neq(0,0,0)$, the vector $(A\,X\,Y)\mathtt{G}(e_x,e_y,L)$ vanishes for at most one angle L.*

Proof. Removing denominators, the equations $A\mathtt{a}_a+X\mathtt{a}_x+Y\mathtt{a}_y=0$ and $X\mathtt{b}_x+Y\mathtt{b}_y=0$ can be written:

$$\begin{aligned}\big(2\,e_xA+2(1-e^2)X\big)\cos L+\big(2\,e_yA+2(1-e^2)Y\big)\sin L &\\ = -(1+e^2)A-2\,e_x(1-e^2)X-2\,e_y(1-e^2)Y&\\ \big(-2\,e_xe_yX+(e_x^2-e_y^2+1)Y\big)\cos L+\big((e_x^2-e_y^2-1)X+2\,e_xe_yY\big)\sin L&\\ = 2\,e_yX-2\,e_xY.&\end{aligned}$$

If $\Delta = \begin{vmatrix} 2\,e_xA+2(1-e^2)X & 2\,e_yA+2(1-e^2)Y\\ -2\,e_xe_yX+(e_x^2-e_y^2+1)Y & (e_x^2-e_y^2-1)X+2\,e_xe_yY\end{vmatrix}$ is nonzero, there is clearly at most one solution L. If $\Delta=0$, there exists $\lambda\neq0$ such that

$$\begin{aligned}2\,e_xA+2(1-e^2)X &= \lambda\left(-2\,e_xe_yX+(e_x^2-e_y^2+1)Y\right),\\ 2\,e_yA+2(1-e^2)Y &= \lambda\left((e_x^2-e_y^2-1)X+2\,e_xe_yY\right),\end{aligned}$$

and there may be a solution to the system above only if

$$(1+e^2)A+2\,e_x(1-e^2)X+2\,e_y(1-e^2)Y=-2\,\lambda(e_yX-e_xY)$$

These three equations forms a linear system in (A, X, Y), $M(A, X, Y)^T = 0$ with

$$M = \left(\begin{array}{ccc} 2e_x & 2(1 - e^2 + \lambda e_x e_y) & -\lambda(e_x^2 - e_y^2 + 1) \\ 2e_y & -\lambda(e_x^2 - e_y^2 - 1) & 2(1 - e^2 - \lambda e_x e_y) \\ (1 + e^2) & 2\left(e_x(1 - e^2) + \lambda e_y\right) & 2\left(e_y(1 - e^2) - \lambda e_x\right) \end{array} \right).$$

A brief computation gives $\det M = (1 - e)^3 (1 + e)^3 (\lambda^2 + 4)$, strictly positive when $0 \le e < 1$. Hence $M(A, X, Y)^T = 0$ implies $(A, X, Y) = 0$. $\square$

Since the rank of $\mathtt{G}$ is obviously equal to 2 and the rank of $\{\mathtt{G}, \partial\mathtt{G}/\partial L\}$ equal to 3 for any (e_x, e_y, L), the hypotheses of Theorem 4.7 are satisfied by the planar control 2-body system, and it guarantees existence of a flow for the Hamiltonian system governing the extremals of minimum time for its average system.

6. Conclusion. We have defined an average control system in a way that differs from the usual use of averaging in control theory, and given its basic properties: an equivalent for control systems of the averaging theorems for ODEs that justifies the average control system as such, and statements on its description and regularity. It has natural applications to approximation of minimum time control. It has already allowed us to give (with restrictions on the eccentricities coming from the point raised in Remark 5.1) a proof that the minimum time between 2 ellipses grows like $1/\varepsilon$ for the planar 2-body problem [6].

The present results are however mostly a starting point. The regularity of H has to be further explored when the conditions of Theorem 3.22 do not hold, see the last paragraph of §3. Explicit computation of the average system and its extremals for the 2-body problem has to be conducted.

Acknowledgements. The authors are indebted to Jana Němcová from Institute of Chemical Technology, Prague, for a careful proof-reading of the draft manuscript, and the two anonymous referees from this journal for extremely constructive reviews that make this paper considerably easier to read than the original submission.

Appendix. Proof of Proposition 3.16. We first state two lemmas.

LEMMA A.1. *Assume that $\bar{X} \in \widetilde{\mathcal{Z}}$ and (39) is satisfied. There is a neighborhood U of $\bar{X}$ in O^{d} and a smooth map $\widehat{\chi} : U \to S^1$ such that, for $(\theta, X) \in U$, one has $\mathsf{V}(\theta, X) = 0$ only if $\theta = \widehat{\chi}(X)$, and*

$$\left(\frac{\partial \mathsf{V}}{\partial \theta}(\widehat{\chi}(X), X) \,\middle|\, \mathsf{V}(\widehat{\chi}(X), X) \right) = 0\,, \quad X \in U\,. \tag{A.1}$$

Proof. From (39.a), $\mathcal{Z} = \{(\theta, X) \in S^1 \times O^{\mathsf{d}}\,,\ \mathsf{V}(\theta, X) = 0\}$ is a smooth submanifold of $S^1 \times O^{\mathsf{d}}$ and from (39.c), $\widetilde{\mathcal{Z}}$ given by (34) a smooth submanifold of O^{d}, both of dimension $d + 1 - m$, and the projection $\pi : S^1 \times O^{\mathsf{d}} \to O^{\mathsf{d}}$ induces a diffeomorphism $\mathcal{Z} \to \widetilde{\mathcal{Z}}$ whose inverse is of the form $x \mapsto (\chi(x), x)$ with χ a smooth map $\widetilde{\mathcal{Z}} \to S^1$ that satisfies, for all $X \in \widetilde{\mathcal{Z}}$: $\mathsf{V}(\theta, X) = 0$ if an only if $\theta = \chi(x)$.

Consider the map $T : S^1 \times O^{\mathsf{d}} \to \mathbb{R}$ given by $T(\theta, X) = \left(\frac{\partial \mathsf{V}}{\partial \theta}(\theta, X) \,\middle|\, \mathsf{V}(\theta, X) \right)$. Let $\bar{X}$ be in $\widetilde{\mathcal{Z}}$; since $\mathsf{V}(\chi(\bar{X}), \bar{X}) = 0$, one has $T(\chi(\bar{X}), \bar{X}) = 0$ and $\partial T/\partial\theta(\chi(\bar{X}), \bar{X}) = \|\frac{\partial \mathsf{V}}{\partial \theta}(\chi(\bar{X}), \bar{X})\|^2$, nonzero from assumption (39.b): the implicit function theorem yields a unique map $\widehat{\chi}$ from a neighborhood U of $\bar{X}$ in O^{d} to a neighborhood of $\chi(\bar{X})$ in S^1 such that $\theta = \widehat{\chi}(X)$ solves $T(\theta, X) = 0$; it must therefore coincide with χ in $U \cap \widetilde{\mathcal{Z}}$ and satisfies the lemma. $\square$

LEMMA A.2. *Assume that $\bar{X} \in \widetilde{\mathcal{Z}}$ and (39) is satisfied. There exist a neighborhood U of $\bar{X}$ in O^{d}, local coordinates $x_1, \ldots, x_d$ defined on U, and smooth maps $P\colon U \to SO(m)$, $\alpha\colon U \to \mathbb{R}$, and $W\colon S^1 \times U \to \mathbb{R}^m$ such that, with $X_{\mathbf{I}} = \begin{pmatrix} x_1 \\ \vdots \\ x_{m-1} \end{pmatrix}$,*

$$\mathsf{V}(\theta, X) = P(X)\Big[\begin{pmatrix} X_{\mathbf{I}} \\ \alpha(X)\ (\theta - \widehat{\chi}(X)) \end{pmatrix} + (\theta - \widehat{\chi}(X))^2\ W(\theta, X)\Big] \tag{A.2}$$

$$= P(X)\Big[\begin{pmatrix} X_{\mathbf{I}} \\ 0 \end{pmatrix} + (\theta - \widehat{\chi}(X))\ W_1(\theta, X)\Big] \tag{A.3}$$

$$\text{with } W_1(\theta, X) = \begin{pmatrix} 0_{m-1} \\ \alpha(X) \end{pmatrix} + (\theta - \widehat{\chi}(X))\ W(\theta, X)\ , \tag{A.4}$$

in $S^1 \times U$, where α is bounded from below: $0 < \alpha_0 < \alpha(X)$, $X \in U$. Furthermore, for a constant $K_3 > 0$, one has, for all $(\theta, X) \in S^1 \times U$,

$$\|\mathsf{V}(\theta, X)\| \geq K_3 \sqrt{\|X_{\mathbf{I}}\|^2 + \alpha(X)^2\,(\theta - \widehat{\chi}(X))^2}\,, \tag{A.5}$$

and

$$X_{\mathbf{I}} = 0\ \Rightarrow \|W_1(\theta, X)\| \geq K_3\,. \tag{A.6}$$

Proof. The map $X \mapsto \frac{\partial \mathsf{V}}{\partial \theta}(\widehat{\chi}(X), X)$ is nonzero for $X = \bar{X}$, hence it does not vanish on a sufficiently small neighborhood U of $\bar{X}$, and one may write

$$\frac{\partial \mathsf{V}}{\partial \theta}(\widehat{\chi}(X), X)\ =\ P(X)\ \begin{pmatrix} 0_{m-1} \\ \alpha(X) \end{pmatrix}\ ,\quad \alpha(X) > \alpha_0 > 0\ . \tag{A.7}$$

Define $v_1, \ldots, v_m$, smooth maps $S^1 \times U \to \mathbb{R}$ by

$$\begin{pmatrix} v_1(\theta, X) \\ \vdots \\ v_m(\theta, X) \end{pmatrix} = P^{-1}(X)\,\mathsf{V}(\theta, X)\,. \tag{A.8}$$

For i between 1 and $m-1$, $\frac{\partial v_i}{\partial \theta}(\widehat{\chi}(X), X) = 0$ from (A.7), and $v_i(\widehat{\chi}(\bar{X}), \bar{X}) = 0$ from Lemma A.1 and, using(39.a), the rank of the map $X \mapsto (v_1(\widehat{\chi}(X), X), \ldots, v_{m-1}(\widehat{\chi}(X), X))$ is $m-1$ at $X = \bar{X}$: on a possibly smaller neighborhood U, there are local coordinates $x_1, \ldots, x_d$ such that $v_i(\theta, X) = x_i + (\theta - \widehat{\chi}(X))^2\ W_i(\theta, X)$ for $i \leq m-1$ and for some smooth W_i; substituting (A.7) and (A.8) in (A.1) implies $v_m(\widehat{\chi}(X), X) = 0$, hence $v_m(\theta, X) = \alpha(X)\ (\theta - \widehat{\chi}(X)) + W_m(\theta, X)\,(\theta - \widehat{\chi}(X))^2$ for a smooth W_m; (A.2) is proved.

Possibly restricting U to a subset with compact closure, $\|W(\theta, X)\|$ is bounded on $S^1 \times U$; if $|\theta - \widehat{\chi}(X)| \leq \frac{1}{2}\alpha_0/\max\|W\|$, then (A.5) holds with $K_3 = \frac{1}{2}$ according to (A.2); on the set where $|\theta - \widehat{\chi}(X)| \geq \frac{1}{2}\alpha_0/\max\|W\|$, V does not vanish and hence $(\|X_{\mathbf{I}}\|^2 + \alpha(X)^2\,(\theta - \widehat{\chi}(X))^2)^{1/2}/\|\mathsf{V}(\theta, X)\|$ is bounded from below; (A.5) is proved, with K_3 smaller than this bound and than $\frac{1}{2}$. From (A.4), $W_1(\widehat{\chi}(\bar{X}), \bar{X}) \neq 0$ because α does not vanish; from assumption (39.b) and (A.3) (where $X_{\mathbf{I}} = 0$ if $X = \bar{X}$), $W_1(\theta, \bar{X}) \neq 0$ if $\theta \neq \widehat{\chi}(\bar{X})$, hence W_1 does not vanish on $S^1 \times \{\bar{X}\}$; it is therefore bounded from below on $S^1 \times U$ with U a small enough neighborhood of $\bar{X}$: (A.6) holds with K_3 smaller than this bound. ∎

Proof of Proposition 3.16. We use $[-\pi,\pi]$ instead of $[0,2\pi]$ as an interval of integration. Let $h\in\mathbb{R}^{\mathsf{d}}$, with $\|h\|=1$. From (36), one has, for some constant $\widetilde{K}$ using bounds on the derivatives of the smooth V,

$$\begin{aligned}|\mathrm{d}\mathsf{H}(X).h-\mathrm{d}\mathsf{H}(Y).h| &\leq \left|\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\partial\mathsf{V}}{\partial X}(\theta,X).h-\frac{\partial\mathsf{V}}{\partial X}(\theta,Y).h\left|\frac{\mathsf{V}(\theta,X)}{\|\mathsf{V}(\theta,X)\|}\right.\right)\mathrm{d}\theta\right|\\ &\quad+\left|\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\partial\mathsf{V}}{\partial X}(\theta,Y).h\left|\frac{\mathsf{V}(\theta,X)}{\|\mathsf{V}(\theta,X)\|}-\frac{\mathsf{V}(\theta,Y)}{\|\mathsf{V}(\theta,Y)\|}\right.\right)\mathrm{d}\theta\right|\\ &\leq \widetilde{K}\|X-Y\|+\frac{\widetilde{K}}{2\pi}\left\|\int_{-\pi}^{\pi}\frac{\mathsf{V}(\theta,X)}{\|\mathsf{V}(\theta,X)\|}\mathrm{d}\theta-\int_{-\pi}^{\pi}\frac{\mathsf{V}(\theta,Y)}{\|\mathsf{V}(\theta,Y)\|}\mathrm{d}\theta\right\|.\end{aligned}$$

Finally, defining

$$\widehat{\mathsf{V}}(\varphi,X)=\mathsf{V}(\widehat{\chi}(X)+\varphi,X)\ ,\quad \widehat{W_1}(\varphi,X)=W_1(\widehat{\chi}(X)+\varphi,X)\ , \tag{A.9}$$

and making a different change of variables in the last two integrals, one has

$$\begin{aligned}\|\mathrm{d}\mathsf{H}(X).h-\mathrm{d}\mathsf{H}(Y).h\| &\leq \widetilde{K}\|X-Y\|+\frac{\widetilde{K}}{2\pi}\int_{-\pi}^{\pi}\left\|\frac{\widehat{\mathsf{V}}(\varphi,X)}{\|\widehat{\mathsf{V}}(\varphi,X)\|}-\frac{\widehat{\mathsf{V}}(\varphi,Y)}{\|\widehat{\mathsf{V}}(\varphi,Y)\|}\right\|\mathrm{d}\varphi\\ &\leq \widetilde{K}\|X-Y\|+\frac{\widetilde{K}}{\pi}\int_{-\pi}^{\pi}\frac{\|\widehat{\mathsf{V}}(\varphi,X)-\widehat{\mathsf{V}}(\varphi,Y)\|}{\|\widehat{\mathsf{V}}(\varphi,X)\|}\,\mathrm{d}\varphi \end{aligned}\tag{A.10}$$

where the last inequality uses the fact $\|\frac{u}{\|u\|}-\frac{v}{\|v\|}\|\leq 2\min\{\frac{\|u-v\|}{\|u\|},\frac{\|u-v\|}{\|v\|}\}$, and also holds with $\|\widehat{\mathsf{V}}(\varphi,Y)\|$ instead of $\|\widehat{\mathsf{V}}(\varphi,X)\|$ in the denominator. Now let us use Lemma A.2, let $X=(x_1,\ldots,x_d)$ and $Y=(y_1,\ldots,y_d)$ in these coordinates; from (A.3), one has, with $\widehat{W_1}$ defined by (A.9),

$$\widehat{\mathsf{V}}(\varphi,X)=P(X)\Big[\begin{pmatrix}X_{\mathbf{I}}\\0\end{pmatrix}+\varphi\,\widehat{W_1}(\varphi,X)\Big],\ \widehat{\mathsf{V}}(\varphi,Y)=P(Y)\Big[\begin{pmatrix}Y_{\mathbf{I}}\\0\end{pmatrix}+\varphi\,\widehat{W_1}(\varphi,Y)\Big]. \tag{A.11}$$

Hence
$$\begin{aligned}\widehat{\mathsf{V}}(\varphi,X)-\widehat{\mathsf{V}}(\varphi,Y) &= \big(P(X)-P(Y)\big)P(X)^{-1}\widehat{\mathsf{V}}(\varphi,X)\\ &\quad+P(Y)\Big[\varphi\Big(W_1(\varphi,X)-W_1(\varphi,Y)\Big)+\begin{pmatrix}X_{\mathbf{I}}-Y_{\mathbf{I}}\\0\end{pmatrix}\Big]\end{aligned}$$

and finally

$$\frac{\|\widehat{\mathsf{V}}(\varphi,X)-\widehat{\mathsf{V}}(\varphi,Y)\|}{\|\widehat{\mathsf{V}}(\varphi,X)\|}\leq\|P(X)-P(Y)\|+\frac{|\varphi|\,\|W_1(\varphi,X)-W_1(\varphi,Y)\|}{\|\widehat{\mathsf{V}}(\varphi,X)\|}+\frac{\|X_{\mathbf{I}}-Y_{\mathbf{I}}\|}{\|\widehat{\mathsf{V}}(\varphi,X)\|}. \tag{A.12}$$

Two cases are to be distinguished:

(i) If $X_{\mathbf{I}}=Y_{\mathbf{I}}=0$, then φ factors out of $\widehat{\mathsf{V}}(\varphi,X)$ and $\widehat{\mathsf{V}}(\varphi,Y)$ in (A.11) and the last term in (A.12) is zero: according to (A.6), the integrand in (A.10) is bounded by

$$\|P(X)-P(Y)\|+\frac{\|\widehat{W_1}(\varphi,X)-\widehat{W_1}(\varphi,Y)\|}{K_3}\ ,$$

and finally $|\mathrm{d}\mathsf{H}(X).h-\mathrm{d}\mathsf{H}(Y).h|\leq K\,\|X-Y\|$ with a constant K that depends only on V, the open set U and the coordinates.

(ii) If $X_{\mathbf{I}}\neq 0$ (or $Y_{\mathbf{I}}\neq 0$, interchanging X and Y), then (A.12), using (A.5), implies that the integrand in (A.10) is bounded by

$$\|P(X)-P(Y)\|+\frac{1}{K_3}\,\frac{1}{\alpha_0}\ \|W_1(\varphi,X)-W_1(\varphi,Y)\|+\frac{1}{K_3}\sqrt{\frac{\|X_{\mathbf{I}}-Y_{\mathbf{I}}\|^2}{\|X_{\mathbf{I}}\|^2+\alpha(X)\varphi^2}}\ .$$

but the same is also true replacing $\alpha(X)$ with $\alpha(Y)$ and $\|X_{\mathbf{I}}\|^2$ with $\|Y_{\mathbf{I}}\|^2$; hence, since $\|a-b\|^2 \leq 4\max\{\|a\|^2, \|b\|^2\}$, the last term may be replaced by $\frac{2}{K_3}\sqrt{\frac{\|X_{\mathbf{I}}-Y_{\mathbf{I}}\|^2}{\|X_{\mathbf{I}}-Y_{\mathbf{I}}\|^2+4\alpha_0\varphi^2}}$, whose integral between $-\pi$ and π is equal to

$$\frac{\|X_{\mathbf{I}}-Y_{\mathbf{I}}\|}{K_3\sqrt{\alpha_0}}\ \ln(1+\frac{4\pi\sqrt{\alpha_0}}{\|X_{\mathbf{I}}-Y_{\mathbf{I}}\|}+\frac{8\pi^2\alpha_0}{\|X_{\mathbf{I}}-Y_{\mathbf{I}}\|^2}),$$

which is less than $\|X_{\mathbf{I}}-Y_{\mathbf{I}}\|(k_1+k_2\ln\frac{1}{\|X_{\mathbf{I}}-Y_{\mathbf{I}}\|})$ for some k_1, k_2 when, say, $\frac{\|X_{\mathbf{I}}-Y_{\mathbf{I}}\|}{2\sqrt{\alpha_0}} < 1$, and finally, since $\|X_{\mathbf{I}}-Y_{\mathbf{I}}\|$ is less than $\|X-Y\|$ and $u \mapsto u\ln(1/u)$ is nondecreasing, less than $\|X-Y\|(k_1+k_2\ \ln\frac{1}{\|X-Y\|})$.

Cases (i) and (ii) do imply (40), possibly restricting U so that $\ln\frac{1}{\|X-Y\|} \geq 1$. $\square$

REFERENCES

[1] A. A. Agrachev and Y. L. Sachkov. *Control theory from the geometric viewpoint*, volume 87 of *Encyclopaedia Math. Sc.* Springer-Verlag, Berlin, 2004.

[2] V. I. Arnold. *Mathematical methods of classical mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2nd edition, 1989.

[3] D. Bao, S.-S. Chern, and Z. Shen. *An introduction to Riemann-Finsler geometry*, volume 200 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2000.

[4] A. Bombrun. *Les Transferts Orbitaux à Faible Poussée : Optimalité et Feedback*. PhD thesis, Ecole des Mines de Paris, Mar. 2007.

[5] A. Bombrun and J.-B. Pomet. On the averaged control system. In *Proceedings of the 17th MTNS*, pages 2912–2917, 2006.

[6] A. Bombrun and J.-B. Pomet. Asymptotic behavior of time optimal orbital transfer for low thrust 2-body control system. *DCDS Supplements*, pages 122–129, 2007.

[7] B. Bonnard and J.-B. Caillau. Riemannian metric of the averaged energy minimization problem in orbital transfer with low thrust. *Annales de l'Institut Henri Poincaré*, 2006.

[8] F. Bullo. Averaging and vibrational control of mechanical systems. *SIAM J. Control Optim.*, 41(2):542–562, 2002.

[9] J.-B. Caillau and J. Noailles. Coplanar control of a satellite around the Earth. *ESAIM Control Optim. Calc. Var.*, 6:239–258 (electronic), 2001.

[10] F. Chaplais. Averaging and deterministic optimal control. *SIAM J. Control Optim.*, 25(3):767–780, 1987.

[11] F. Clarke, Y. Ledyaev, R. Stern, and P. Wolenski. *Nonsmooth Analysis and Control Theory*, volume 178 of *Graduate Texts in Mathematics*. Springer, 1998.

[12] M. Fliess, J. Lévine, P. Martin, and P. Rouchon. Flatness and defect of nonlinear systems: Introductory theory and examples. *Internat. J. Control*, 61(6):1327–1361, 1995.

[13] S. Geffroy. *Généralisation des techniques de moyennation en contrôle optimal - Application aux problèmes de rendez-vous orbitaux en poussée faible*. Thèse de doctorat, Institut National Polytechnique de Toulouse, Toulouse, France, Oct. 1997.

[14] S. Geffroy and R. Epenoy. Optimal low-thrust transfers with constraints-generalization of averaging technics. *Acta Astronautica*, 41(3):133–149, 1997.

[15] P. Hartman. *Ordinary differential equations*. Birkhauser, 1982. 2nd edition.

[16] T. Kailath. *Linear systems*. Information and System Sciences. Prentice-Hall Inc., Englewood Cliffs, N.J., 1980.

[17] J. Kurzweil and J. Jarnik. Iterated Lie brackets in limit processes in ordinary differential equations. *Results in Mathematics*, 14:125–137, 1988.

[18] W. Liu. Averaging theorems for highly oscillatory differential equations and iterated Lie brackets. *SIAM J. Control Optim.*, 35(6):1989–2020, Nov. 1997.

[19] S. M. Meerkov. Principle of vibrational control: theory and applications. *IEEE Trans. Automat. Control*, 25(4):755–762, 1980.

[20] P. Morin, J.-B. Pomet, and C. Samson. Design of homogeneous time-varying stabilizing control laws for driftless controllable systems via oscillatory approximation of lie brackets in closed loop. *SIAM J. Control Optim.*, 38(1):22–49, 1999.

[21] R. M. Murray. Nilpotent bases for a class of nonintegrable distributions with applications to trajectory generation for nonholonomic systems. *Math. of Control, Signals & Systems*, 7:58–75, 1994.

[22] L. S. Pontryagin, V. G. Boltjanskiĭ, R. V. Gamkrelidze, and E. Mitchenko. *Théorie mathématique des processus optimaux*. Editions MIR, Moscou, 1974.

[23] J. A. Sanders and F. Verhulst. *Averaging Methods in Nonlinear Dynamical Systems*, volume 56 of *Applied Mathematical Sciences*. Springer-Verlag, 1985.

[24] R. Schneider. *Convex bodies: the Brunn-Minkowski theory*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1993.